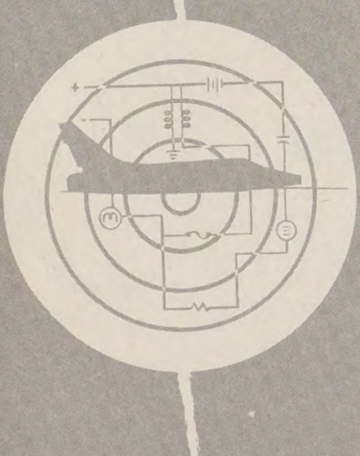
AF MANUAL 136-25

Armament Electronic Systems (Interceptor)

DEPARTMENT OF THE AIR FORCE

AF MANUAL 136-25

Armament Electronic Systems (Interceptor)



DEPARTMENT OF THE AIR FORCE

AIR FORCE MANUAL
NUMBER 136-25

DEPARTMENT OF THE AIR FORCE
WASHINGTON 25, D. C.
28 DECEMBER 1959

FOREWORD

This Manual has been prepared for use as an adjunct to the training of officers and airmen engaged in the maintenance of interceptor armament electronic systems in the United States Air Force.

An understanding of basic electronics, particularly vacuum tube theory, is assumed. The topics discussed cover in a general manner all the principal components of fire control and weapons control systems that can be presented in an unclassified publication. Despite the limitations imposed by security considerations, the scope of this Manual is so broad that many subordinate topics cannot be afforded an exhaustive treatment. In such cases, coverage is confined to representative examples or circuits.

This Manual begins with a discussion of National Air Defense and the role of the interceptor therein. Then, after a brief historical development of the interceptor aircraft and its armament, the fire control problem is analyzed and discussed in detail. This is followed by chapters discussing the components of typical fire control and weapons control systems. The Manual concludes with a detailed analysis and discussion of the harmonization problem as applied to interceptor aircraft.

Recommendations for the improvement of this Manual are encouraged and should be forwarded to Headquarters Air Training Command, Randolph Air Force Base, Texas.

BY ORDER OF THE SECRETARY OF THE AIR FORCE:

OFFICIAL:

THOMAS D. WHITE
Chief of Staff

J. L. TARR
Colonel, USAF
Director of Administrative Services

OPI: AFPTR
DISTRIBUTION: F

This training manual is published by the Commander, Air Training Command, in accordance with provisions of AFR 5-39

CONTENTS

CONTENTS—Continued

CHAPTER 1

AIR-TO-AIR INTERCEPTION

The opening section of this chapter, The Air Defense of the United States and the Role of the Interceptor, was written well in advance of the initial launchings of earth satellites both here and in the Soviet Union. The military significance of these successes has been the subject of many controversial articles in our newspapers and magazines. Attention was focused not so much on the tiny moons themselves, but on the rocket vehicles used to place them in their orbits. Are such vehicles but a step away from a successful intercontinental ballistic missile? We can form our own answer to this question from the fact that at least one missile program was put on a "crash" basis almost immediately following the highly publicized failure of our first attempt to launch a similar satellite. Shortly afterward, the head of a concern engaged in developing a weapon of that type was quoted as saying that before long we would have a successful ICBM.

The concepts of strategic bombing and air defense are in the process of undergoing revolutionary changes. It is not within the scope of this manual to discuss the nature of these changes, nor to outline their present and future developments. We must leave the dissemination of specific information to official channels, retaining our discussion of air defense as it stands even though the concepts upon which it is based may be rendered obsolete within the next year or two—if not already.

AIR DEFENSE

The range of modern military aircraft has been increased to the point where vital areas of no nation on earth are safe from air attack by reason of distance alone. And along with the development of the means of large scale intercontinental air operations, considerable progress has been made in increasing the violence and efficiency of mass-destruction weapons that are deliverable from the air.

For a nation such as the United States, because of its highly organized industrial society and many great population centers, the mere thought of a surprise attack by an aggressor having weapons of devastating power produces extremely critical implications. Consequently, forces-in-being for defense are indispensable to our national security. Of these, the Air Force provides forces and measures for air defense to preserve the nation's vital resources.

The primary object of air defense is to insure national survival. However, purely defensive measures have limitations which make them incapable of insuring immunity from air attack. The military policy for the

defense of the United States recognizes that, in the event of attack, air defense measures coupled with strong air counterblows against the aggressor's strength will provide the best security.

The Air Defense of the United States

The immediate basic objective of all forces-in-being is to discourage any enemy from attacking the United States. However, the possibility of a rash attack must not be overlooked. In this case the primary objectives of air defense are: (1) to prevent hostile air forces from inflicting such critical damage that national survival is endangered; (2) to preserve the means for conducting sustained defensive and offensive operations by air to defeat the enemy; and (3) to aid in sustaining our Nation's will to resist the enemy.

Interdependence of Defensive and Offensive Forces

A ready air defense system possesses the means to destroy attacking air forces or to reduce the effectiveness of their attacks. As mentioned previously, purely defensive measures have very definite limitations. These limitations make it necessary to launch an immediate air counteroffensive whose primary objective is to destroy the enemy's air capabilities at the source before their full strength can be brought to bear against the United States in sustained operations. Therefore, the long-range air forces which are needed to conduct a powerful strategic counteroffensive must—like the air defense forces —be available at the *outset* of hostilities.

The relative importance of the responsibilities for air defense against hostile air forces and for offensive action against the sources of enemy attack are such that neither force can be increased to the point of denying essential strength to the other. While offensive forces act to reduce the capability of the enemy for sustained offensive action, they cannot take the place of air defense forces designed and deployed for the immediate protection of the United States. On the other hand, although air defense aids the offensive effort by destroying hostile air forces engaged in the attack, it cannot replace offensive air forces. Hence, our defensive and offensive forces must be mutually supporting and operate with a singleness of purpose.

THE AIR-DEFENSE TEAM. Active air defense is the responsibility of the Air Force and units of the other military services, but we can carry that responsibility a step further. It is the responsibility of the air defense team. This team is made up of three principal elements — interceptor aircraft, antiaircraft artillery (AAA), and aircraft control and warning (AC&W) units.

Interceptor aircraft. Interceptor aircraft engage and destroy enemy aircraft in the air—preferably before they can complete their bombing missions. The term *interceptor* may be applied to any aircraft used to intercept hostile aircraft even though not specifically designed for air-defense operations.

Antiaircraft Artillery. Antiaircraft-artillery units complement interceptors in their mission. Heavy and medium guns, automatic weapons, and ground-to-air missiles sited near a target area attempt to engage and destroy as many hostile air targets as are not at the moment engaged by interceptors. In certain cases, simultaneous engagement by interceptors and AAA may be undertaken.

In our defense system, AAA units are regularly assigned to the Army Antiaircraft Command (ARAACOM). In actual combat, however, the USAF Air Defense Command (ADC) exercises operational (fire) control over such units. Both ARAACOM and ADC are under the Continental Air Defense Command (CONAD).

Aircraft Control and Warning Units. Aircraft control and warning units are part of ADC and constitute a system of control centers, radar stations, and communications facilities. This system detects and identifies

aircraft in flight, controls the action of interceptor aircraft against hostile targets, and also controls the firing of AAA units against hostile targets. In continental United States, AC&W units are assigned to each air division (defense). AC&W squadrons within each division operate radar stations and control centers. The Ground Observer Corps (GOC), an organization of volunteer civilian aircraft spotters, complements AC&W radar coverage with visual aircraft detection.

IMPLEMENTING AIR DEFENSE. The air-defense team, outlined briefly above, accomplishes the air-defense mission in four broad steps: (1) detection of approaching aircraft; (2) identification of them as friend or foe; (3) interception of hostile aircraft; and (4) destruction of such aircraft either by interceptor aircraft or AAA fire.

The requirements for conducting and accomplishing the above four steps include: (1) advance information on hostile and friendly air activity; (2) the means of identifying aircraft; (3) interceptor and AAA units—ready and suitably deployed; (4) a system to control interceptors and to coordinate them with ground-based fire; and (5) a system of air-defense bases placed to make maximum use of available air-defense forces against enemy aircraft. In addition, there is a requirement for long-range intelligence on the enemy's disposition and his capability for launching an air attack.

The Air Force is entirely or partially responsible for all four steps taken by the air-defense team to implement air defense. The Air Force provides most of the radars involved in detecting and identifying approaching aircraft and, in most instances, all of the interceptors used to combat hostile forces in the air. Furthermore, the Air Force has fire control over AAA units in areas where interception activity is in progress. Aside from the actual destruction of hostile air targets accomplished by interceptor squadrons, Air Force measures taken to implement air defense are handled almost entirely by AC&W units. They man the AF control and direction centers, the ground-controlled interception (GCI) stations, and early-warning radars. They also supply the controllers who track incoming targets, establish interception approaches, and direct interceptors in for the kill.

THE ROLE OF THE INTERCEPTOR

Interceptor units, in place and prepared for operations, are capable of exerting significant deterrent pressure against the enemy even before an attack occurs. The capability for effective defense confronts the enemy with the immediate threat of high losses in his attacking forces. This imposes certain restrictive conditions upon the enemy and, to a degree, limits the courses of action that his forces can employ against the defended territory. Moreover, the interceptor force-in-being greatly reduces the enemy's chances of achieving an initial advantage through surprise in the first attack and may cause so many complications in his operations that major failure will occur.

Functions of Interceptor Units

Interceptor units exploit their capabilities to the utmost to provide security from enemy attack by air. It is recognized that maximum security requires the *elimination of the enemy's total power to attack by air* and that it is never complete until hostile air forces—wherever they may be—cease to exist as an *organized, effective fighting force.*

The interceptor mission is to destroy invading air vehicles, to impose tactical restrictions upon their employment, and—when attacks are delivered—to minimize their effects. These activities do not imply that security from air attack is solely a matter of defense. Neither do they imply that security can be dealt with altogether in terms of limited geographical areas. Although it is essentially concerned with the defense of the United States, this mission must give due consideration to our offensive objectives and the strategy of global conflict.

The foregoing is reflected in the interceptor functions, which are:

a. To defend areas essential to the prosecution of the war

b. To deny freedom of action to hostile forces

c. To destroy, divert, or deceive attacking forces

d. To minimize the effects of enemy attacks

e. To participate in gaining command of the air

f. To participate in other action as directed (after command of the air has been achieved) to complete the defeat of the enemy.

Required Interceptor Capabilities

The nature of enemy attack capabilities makes it mandatory that interceptor forces possess certain capabilities that we will discuss briefly.

CONSTANT HIGH DEGREE OF READINESS. Interceptor forces must be prepared to function under unfavorable and difficult conditions imposed by the enemy in exercising the initiative of attack. They must have the capability to maintain, over a long period of time, varying operations ranging from standby conditions to full alert.

SPEED OF ACTION. These forces must have the capability to alert and energize themselves. When command decisions have been made, they must perform their functions and tasks with exceptional speed because only limited time will be available once the attacking enemy forces have been detected.

ENDURANCE. Interceptor forces must be capable of sustaining their operations for all periods of time during which the enemy is capable of conducting air attacks.

DAY, NIGHT, INCLEMENT WEATHER, AND FAIR WEATHER OPERATIONAL CAPABILITY. Since the enemy has the option of attacking during any or all such conditions, interceptor forces must be proportioned as to specialized weapons capabilities so that the total force includes the ability to operate under all these conditions.

CAPABILITY FOR OTHER OPERATIONS. Interceptor forces should have characteristics which permit them to be employed in other operations against the enemy to hasten the conclusion of the war.

Employment of Interceptors

Interceptors may be employed in either general or local defense. The speed and range of an interceptor aircraft enables it to provide protection for a group of objectives spread over a wide area or, with equal facility, to restrict its operation to one of local defense or the protection of specific objectives. Antiaircraft artillery, since it is not tactically mobile, is unsuited for carrying out a general defense mission. To the contrary, it is capable of providing protection only at a particular location. In view of this fact and, because the amount of available AAA is always limited, many profitable targets will be undefended by AAA. Hence, it may become necessary to employ interceptor aircraft in general defense to provide protection for some of these targets.

DEVELOPMENT OF INTERCEPTOR AIRCRAFT AND ARMAMENT

In tracing the history of interceptor aircraft, we will consider three periods in military aviation; the early years, 1914-1922; the modern era, 1922-1945; and the jet age, 1945 to date. As in the case of most aspects of human history, no clean-cut lines of demarcation separate these periods. We must expect a reasonable degree of overlapping.

Due to the limitations of time and space, our history cannot go much beyond the mere outlining of the development of interceptors and their armament. Since 1909, the USAF and its predecessors have operated, tested, or seriously considered well over 2,000 models

and modifications of the various types of military and utility aircraft. A fair percentage of these were fighter/interceptors.

Through the years only a comparatively few models have been specifically designed and operated exclusively as interceptors. To present a more complete and well-rounded story, we have chosen to encompass the entire fighter field. Except when important developments occurred abroad, our historical discussion will be confined to aircraft of the United States. With reference to our own aircraft, dates consisting only of the year generally refer to the *fiscal year* in which initial delivery of the specific model was made.

The Early Years, 1914-1922

Too many people are inclined to ignore or discount the importance of World War I (1914-1918) with respect to the advancement of military aviation. They do not realize that prior to World War I military aviation as we know it today did not exist. This brief span of four years witnessed not only the birth of military aviation itself but the development of virtually all the basic types of military aircraft—fighters, night fighters, interceptors, bombers, attack (ground cooperation), observation, and photo-reconnaissance. Furthermore, the advancement of aerial armament during this period was so rapid that, by the end of the war in 1918, most of the basic weapons existing today had passed through at least the early stages of experimentation. This is no exaggeration; the record speaks for itself. Such things as air-to-air rocketry, aerial cannon, and multiple-gun installations were tried out and used with varying degrees of success over 40 years ago. And if we care to look more closely we will even see an early beginning in the field of guided missiles.

WORLD WAR I. The major contending powers—Britain, France, and Germany—entered the war in August of 1914 with few available first-class military aircraft. Despite the fact that aerial weapon-firing and bomb-dropping had been demonstrated both here and abroad in prewar years, early military aircraft were completely unarmed. The high commands of that time did not consider the airplane to be a machine of war. We who have the gift of hindsight must be careful not to be hasty in criticizing them for their low opinion of the airplane. They were composed of practical men who were quick to see that the early airplane lacked ruggedness and dependability, two qualities that are essential in machines of war. The earliest military aircraft put themselves so far "behind the eight ball" that they had to prove themselves under combat conditions before they could be accepted. That they came to be accepted is indicated by the rapid growth and development of military aviation throughout the war. The lack of foresight that hindered the development of air power was evidenced in postwar years *after* the potential of military aircraft had been demonstrated.

The Final Months of 1914. The only picture we have of military aviation in the opening months of the war is one of utter confusion. The first military aircraft were assigned exclusively to reconnaissance and artillery spotting missions. To this there was considerable opposition by the cavalry which considered the field of reconnaissance its exclusive property. Opinions differ considerably as to the effectiveness of early aerial reconnaissance. On one hand we are told that it was a miserable failure at first because the early aircraft did not afford an adequate downward field of view and because inexperience led to the improper evaluation of much aerial intelligence. This was soon corrected by the introduction of improved two-place observation aircraft. On the other hand there are those who contend that aerial observation was successful from its very beginning. In any event, the early air forces took over the job of reconnaissance from the cavalry and started that military branch on its way to oblivion.

The most confusing aspect of this phase of the war is found in the armament of aircraft. First we see military aircraft completely unarmed and observe enemy airmen exchanging friendly greetings when their aircraft pass close by. Then suddenly these same airmen begin using brickbats, monkey wrenches, hand grenades, pistols, carbines, rifles, and even machine guns as aerial weapons.

How did it all begin? Well, no one seems to know for sure. According to tradition, the opening step was taken when an unidentified airman—for reasons known only to himself—thumbed his nose instead of rendering the customary friendly wave of the hand. The offended party reacted to this insult by throwing the nearest loose object at hand, a brickbat or a monkey wrench whose normal function was to keep reconnaissance maps from blowing out of the open cockpit. The immediate result was that airmen began to wear their side arms aloft. When pistol fire proved effective only to the extent of providing relief from the monotony of reconnaissance flying, observers began to provide themselves with carbines and rifles. And when these weapons were found to be only slightly more effective, the ground machine gun was adapted to aerial use.

The principal fault with this story is that it presents a much too orderly development of aerial armament. All evidence points to the fact that the early airmen took upon themselves the arming of their aircraft and that no official action was taken until a common practice had been established. What is more important, the development of armament followed no logical sequence. The machine gun, which became the ultimate weapon of World War I, was among the first weapons to be carried aloft. The war had hardly begun when a British pilot obtained a machine gun and mounted it on his aircraft. Soon afterward he went up in pursuit of an unarmed German *Taube* that was circling nonchalantly over the British airfield. Unfortunately he had not reckoned on the effect of the weight of the gun on the performance of his aircraft. No matter what he did, he could not coax his aircraft above an altitude of 3,500 feet. The enemy plane escaped unharmed, needless to say. And the ingenious airman had his ears blistered by an irate commanding officer because of the unauthorized armament.

The first aerial victory occurred on 26 August 1914 without a single shot being fired. On that date three unarmed British aircraft forced down a similarly unarmed German aircraft by the simple tactics of taking turns making a succession of near-miss passes at it. Soon other British pilots adopted this bloodless method of aerial warfare with success. One cannot blame aviators for choosing capture to the consequences of a mid-air collision. Throughout the war, only balloonists were provided with parachutes; no one seems to have taken the trouble to develop a similar device compact enough to be carried and used by the occupants of combat aircraft.

Several months later we note an order of the German General Staff that provides for the arming of aviators on reconnaissance missions with "pistols and hand grenades." The chances of destroying or disabling an enemy aircraft by using a hand grenade as an air-to-air missile were quite small, but those were days when everyone seemed willing to try anything.

In the early months of the war, the mortality rate for airmen and aircraft was high. However, few losses were caused by enemy action. Unreliable engines, poor maintenance, and the use of hastily converted cow pastures as airfields accounted for the bulk of the casualties. Official German records for the first three months of the war show the loss of over 50 airmen and 100 aircraft. These losses occurred during a period when aircraft was virtually unarmed and antiaircraft artillery was practically nonexistent.

The concept of preventive maintenance with its frequent checking and overhauling was still a long way off. Aircraft were flown until failure of engine, wings, or airframe through sheer exhaustion resulted in crashes.

that adequate maintenance would have eliminated. Then, too, pilots were sent to the front after only a few hours of solo flying. Their inexperience led to frequent crashes in take-off and landing.

The Year 1915. The fighter aircraft came into being early in the year 1915. We should take note, in passing, that two principal factors stood in the way of the evolution of the fighter. First, the early machine gun of World War I was both inaccurate and unreliable. Second, the early engines were so weak that even the burden of a light machine gun seriously affected the performance of the aircraft. Evidence exists that some pilots preferred their pistols rather than the machine gun as an aerial weapon for either or both of these reasons. Hence, true aerial armament had to await the improvement of weapons, engines, and aircraft.

The combination of a dare-devil French aviator and a routine engine malfunction is responsible for the birth of the fighter aircraft. Roland Garros was the Frenchman and the only combat pilot who achieved fame as a flier before the war. Among his exploits were the first crossing of the Mediterranean Sea and several early altitude records. He mounted a light machine gun on the side of the fuselage of his small one-place aircraft; this position was chosen so that he could easily reload the gun from his seat in the cockpit. And he fastened slender pyramid-shaped blocks of steel to the backs of his propeller blades where they passed through the line of fire. In February of 1915, Garros and his aircraft headed into a flight of four German observation planes. He shot down two of them before their crews realized what was happening. The two surviving aircraft high-tailed for home, bringing with them news that demoralized the German air force —the aircraft had become a lethal weapon. Garros became the first of all aces, scoring five victories in 18 days.

Then Garros' aircraft was forced down behind the enemy lines by a clogged fuel line. The aircraft was captured before he was able to destroy it, and the secret of his success was acquired by the enemy. The German High Command promptly presented the Garros propeller to Anthony Fokker, the aircraft designer, and commissioned him to construct a number of aircraft making use of this crude device or one similar to it.

Fokker ground-tested the device on one of his aircraft and quickly concluded that it was not the best solution to the problem of firing a machine gun through an aircraft propeller. Although only about 7% of the bullets fired struck the blocks, each time they did so the resulting vibrations threatened the existence of the propeller and the security of the engine mounts. Further, there was no predicting which way the bullets would be deflected by the blocks. Sooner or later, one would find its way toward the pilot or the engine. In short, the Garros device was unsafe, and only dare-devils like Garros would care to use it.

Fokker decided that the correct solution was to gear the machine gun to the engine and to time the firing of the gun so that bullets passed only between the rotating propeller blades. There are conflicting stories on how Fokker got the basic idea for his synchronizing gear. One version is that it came from his recollection of his boyhood in Holland when one of his pastimes was throwing stones at rotating windmill vanes. The other is that he borrowed the basic idea from a patent that had been filed in Germany several years before the war. Regardless of its source, a successful synchronizing gear was produced and test-fired within 48 hours after Fokker returned to his factory.

This gear in conjunction with a single machine gun was first installed on the Fokker E-1 *Eindekker* (monoplane) fighter. When reports from the front indicated that the weight of the armament system affected the performance of the aircraft, the E-1 was given a more powerful engine and redesignated E-2. This was followed by the E-3, the first aircraft equipped with two synchronized machine guns.

The last of Fokker's early monoplanes, the E-4, was an experimental model equipped

with three synchronized machine guns. Because the E-4 came close to killing one of Germany's foremost aces in an aerial test-firing mission, it was never placed in production. The recoil of three guns was apparently just too much for contemporary airframes and engine mounts to withstand. Hence, two caliber .30 (or its metric equivalent) synchronized guns became standard armament for fighter aircraft, and the practice continued until the approach of World War II.

It is interesting to note that Fokker's early success with the monoplane fighter was a rather exceptional case. Back in 1913 a series of accidents involving single-winged aircraft had resulted in the deaths of several noted European fliers. The British went so far as to ban further use of the monoplane as a training aircraft. This had much to do with the ending of a rapidly growing trend toward the monoplane and the spreading of prejudice against the monoplane throughout Europe. Very few monoplane fighters saw service in World War I. Years would pass before this prejudice among both pilots and officials could be overcome.

The "Fokker Scourge" practically drove the Allied air forces from the skies in the early months of 1915, but for some unknown reason the Germans were never able to take full advantage of the superiority that synchronized guns gave them. The British and the French frantically took up the development of their own synchronizing gear. By late summer they were successful, and the equipping of their aircraft with synchronized guns neutralized the tremendous advantage held by the Germans.

Meanwhile, as an interim measure, various Allied aircraft such as the Martinsyde of the British and the Nieuport Scout of the French were equipped with a single free-firing machine gun mounted on the upper wing to fire over the propeller arc. The Lewis guns used for this purpose were fed by 47-round circular magazines known to airmen as "pie plates." Replacement of these magazines when empty required the pilot to stand up in the cockpit and hold the control stick between his knees. A later version of the over-the-wing gun mount included a curved track so that the gun could be pulled rearward and downward toward the cockpit for reloading purposes.

Other interim measures are both interesting and amusing in retrospect. An early two-place Spad placed a gunner and a movable machine gun in a tiny cockpit located immediately in front of the propeller. It is difficult to think of a more clumsy arrangement, but it seems to have worked. During this period a British aircraft was fitted with a machine gun mounted on the side of the fuselage behind the propeller. However, the gun was angled out so that its line of fire passed outside the propeller arc. Consider the fire control problem presented by this arrangement. The pilot had to fly his aircraft in one direction and aim the gun in another direction. This period also saw the continuance in use of various obsolete pusher-type fighter models whose unrestricted field of fire from a gunner's cockpit in the nose made synchronization unnecessary.

The coming of synchronized guns to both sides initiated true aerial warfare for the first time. This quickly came to be known as "dog fighting," a process of maneuvering one's fighter into a position of advantage behind one's opponent and blasting him at virtual point-blank range. In 1915 Immelmann, the great German ace whose name is perpetuated in the aerial maneuver he invented, taught young pilots in his organization that the maximum effective range in a "dog fight" was approximately 75 yards.

The Year 1916. In this year some Nieuport Scouts had their outer struts fitted to carry Le Prier air-to-air incendiary rockets which were fired electrically. Use of these rockets was limited solely to bringing down enemy observation balloons, more commonly called *blimps.* Their maximum effective range seems to have been less than 400 feet.

Late in 1916 we note the beginning of real progress in fighter aircraft design. Of particular interest among the newly introduced

fighters were the Albatross D-1 of the Germans, the SE-5 of the British, and the Spad 7 of the French. This Spad was the first of a series which collectively became the most popular fighter of World War I here in this country. This popularity was largely based on sentiment; most of our fighter squadrons of World War I were equipped with various models of the Spad.

The Year 1917. Early in 1917 the British introduced what many consider to have been the most powerful fighter of World War I —the Bristol F2B. The "Brisfit" combined the ruggedness of a two-place reconnaissance plane with the speed and maneuverability of a one-place fighter. In addition to one or two synchronized guns, it was armed with twin flexible guns in a cockpit located immediately aft of that of the pilot. Prior to the F2B, common practice in the design of two-seaters placed the gunner's cockpit well aft of the pilot, making teamwork between the pilot and gunner practically impossible. Not only was the Bristol fighter more than a match for any one-place fighter, it is known to have been used as a single-seater by several Allied pilots.

About the same time, the Germans placed into operation the Fokker Dr-1 triplane. The amazing rate of climb and exceptional maneuverability made the "Tripe" what should have been the outstanding interceptor of the war. However, the Fokker triplane was wasted by the German High Command who, in an effort to wrest air superiority away from the Allies, ordered its use as a purely offensive weapon. As such, it was inferior to other German fighters because of its limited fuel capacity and its inability to withstand the effects of a long dive.

Perhaps the best all-around German fighter of early 1917 was the Albatross D-3. The smooth clean fuselage lines of this and later models of the Albatross made it look years ahead of its time. This appearance was created by covering the fuselage with three crossed layers of thin wood veneer. (We used this type of fuselage covering years later just before the advent of the all-metal fighter. A notable example is the YP-24 of 1932.) Similar streamlining appears in other German one-place fighters of 1917 and 1918, particularly those made by Pfalz and Roland.

The Fokker Dr-1 was not the only multiwing aircraft to make its debut in World War I. Triplanes were also produced by Sopwith in Britain and Nieuport in France, but neither of these fighters achieved any great degree of success. Here in America we were not able to escape the fascination of multiwinged aircraft, and prior to our entry into the war several Curtiss S-3 triplane fighters were procured. Nor were three wings the limit. In 1916 the British came out with a quadraplane observation aircraft whose four narrow wings gave it the appearance of a flying venetian blind. The idea behind multiplane aircraft was to give better visibility by dividing the total wing area into a greater number of wings having a more narrow chord than would be possible in a biplane.

In 1917 the British introduced the SE-5A, an improved version of the SE-5. The SE-5A provided the most stable gun platform of any aircraft built in World War I, and yet it was capable of extreme maneuverability. In the SE-5A we see an outstanding example of fighter armament combining both the synchronized and the free-firing machine gun. Its arrangement of one or two synchronized guns on the fuselage and a free-firing gun over the wing points out a disadvantage of synchronization. Since the synchronizing gear stops the firing of a gun whenever a propeller blade approaches the line of fire, the firing rate of the gun is reduced. Furthermore, the firing rate decreases as the engine speed increases, because the propeller blades pass through the line of fire more frequently. Since the maximum firing rate of World War I machine guns was only about 600 rounds per minute, there was a limit to how much interruption of firing could be tolerated without a drastic reduction in fire

power. This was especially true of the SE-5A, which was among the first aircraft to be equipped with a four-bladed propeller.

The cannon-equipped Spad 12 appeared about this time, also. In this aircraft a 37-mm cannon was mounted so as to fire through a hollow propeller shaft. Several French and American pilots operated the Spad 12 successfully. However, it took a good man to do it. The slow firing rate of the cannon made it imperative to score a hit with the first shot. Because the weight of the cannon and its ammunition did not enhance the performance of the aircraft, a miss would enable the intended victim to double back with his more maneuverable fighter and "clobber" the cannoneer with machine-gun fire. Mechanical considerations such as excessive recoil force and poor gun mounts led to the early abandonment of the cannon as a fighter weapon. Yet a number of victories were scored before the Germans even realized that the Allies were using shell-firing guns in their fighters.

The most effective fighter of World War I, the Sopwith Camel, was introduced to combat in July 1917. The Camel downed in aerial combat a total of almost 1,200 enemy aircraft, the greatest number of victories scored with any fighter in World War I. Extremely maneuverable but very tricky to fly, the Camel seems to have been the first single-seater to be modified to a two-seater trainer for the purpose of checking-out pilots. This practice has become rather common in recent years.

In October of 1917 the Junkers D-1 made its first flight. This aircraft — a low-wing monoplane fighter of all-metal construction —was truly years ahead of its time. Prejudice against the monoplane and lack of faith in all-metal construction are the only two reasons that can be given for failure to accept this revolutionary aircraft. A dozen or so years later the soundness of riveting corrugated aluminum sheet to a tubular aluminum structure would be proven in commercial transport aircraft both here and abroad.

We have passed through the year 1917 without mentioning the entry of the United States into the war in April of that year. This was no mere oversight. Our country came into the war without a single up-to-date design in military aircraft. We had an excellent opportunity to observe the advancement of aircraft design through over two years of the war and failed to take advantage of it. This was largely due to the fact that practically no developmental work was sponsored by the government and the aircraft companies had little money to spend on it.

The record of our aircraft industry during World War I can be dismissed with the single word, *disgraceful.* Plans to construct thousands of military aircraft were eventually abandoned in favor of purchasing them from Britain, France, and even Italy. Among the cancelled contracts were orders for 6,000 Spads and 2,000 Bristol fighters. (The Spads were cancelled because the single-seater fighter was "obsolete"; the Bristols because they were "unsafe.") Outside of trainers such as the "Jennie", the only aircraft produced in quantity was the de Havilland DH-4. The DH-4 was supposed to be a two-seater fighter; but, as such, it was obsolete before we began to build them. Those that got overseas had to be rebuilt before they could be flown and saw service as observation and light bombing aircraft. The poor design of the DH-4 won for it the nickname, "Flaming Coffin."

At the end of the war we had 18 or 19 (our sources disagree on the exact number) fighter squadrons in operation at the front. Sixteen of these were equipped with the Spad 13, one with the SE-5A, and one or two with the Sopwith Camel.

Through the period of our participation in the war the engineering division of our air force considered well over a hundred aircraft designs, most of them submitted by the various aircraft companies. About two dozen of them got as far as the experimental flying stage, and several fighter designs were placed in production but too late for wartime service.

The Year 1918. In the summer of 1918 the Fokker D-7 became the standard fighter of the German air force. Many authorities consider the D-7 to have been the best fighter of the war. In the few remaining months of the war it piled up a number of victories that seems to be second only to that of the Sopwith Camel. To be more specific, the D-7 scored 585 victories in *one* month. It is significant to note that the peace treaty that ended the war specified that all remaining Fokker D-7's of the 4,000 that were built be turned over to the Allies. We acquired almost 200 of them, but seem to have destroyed all but a few that were used for experimental purposes.

Just before the end of the war the Fokker D-8, a parasol monoplane fighter, made its appearance. The wing was of full cantilever construction and covered with plywood, requiring no external wire bracing. Official interference with the design of the wing had delayed production for about a year. A few D-8's saw combat service and were promptly dubbed the "Flying Razor" by Allied airmen.

THE POSTWAR YEARS. The war was so expensive that the ensuing period of retrenchment in military spending was inevitable. The years immediately following the war produced little advancement in the design of fighter aircraft. Those left over from the war were continued in operation until they were no longer airworthy. Maintenance funds were diverted to the rebuilding and improvement of World War I aircraft. Most of the effort was concentrated on raising the ceiling of military aircraft through the development of superchargers and improved engines.

To give a more specific idea of how low our air power was permitted to drop after World War I, our entire operational air arm around 1920 consisted of *four* groups, one each: pursuit (fighter), reconnaissance, surveillance, and bombardment. The fighter group was equipped with rebuilt SE-5A's; the remaining groups operated various modifications of the DH-4.

Included among our fighter aircraft of this period we find such aircraft as improved versions of the Fokker D-7 and D-8. What few new aircraft we produced show a very strong influence by the Spad 13. The similarity is especially strong in the case of the Standard E-1 of 1918 (too late for the war), the Lewis and Vought VE-8 of 1919, the Thomas-Morse MB-3 of 1919, the Curtiss-built Orenco D of 1920, and the Boeing MB-3A of 1922 to mention a few. The Orenco D, except for its more powerful engine, was an almost exact duplicate of the Spad design. It is interesting to note that increasing the engine horsepower from 235 to 300 raised the top speed only from 135 to 140 miles per hour.

The Boeing MB-3A, whose design was based on the Thomas-Morse MB-3 which in turn was based on the Spad 13, became the first postwar fighter to be produced in quantity. But the MB-3A's, whose top speed was 141 miles per hour, had hardly begun to roll off the production line in 1922 when an important development in design ushered in the modern era.

The Modern Era, 1922-1945

In 1922 the two Curtiss R-6 racers captured both first and second places in the Pulitzer Race of that year. Not only were they the first aircraft to exceed 200 miles per hour; their success ended the domination of the Spad design with its double bays of wing struts on either side of the fuselage. This racer gave our designers the idea that satisfactory biplanes could be built with less struts and wire bracing than a Spad. From it was developed the Curtiss XPW-8 of 1923, which was paralleled by the similar Boeing XPW-9 of the same year. Both models featured the N-strut originated by Fokker in his D-7 of 1918 and which now became standard for practically all biplanes.

The XPW-8 led to the development of the P-1 of 1925, the first of a long line of Curtiss Hawks. In 1926 a P-1 was fitted with a radial engine and designated XP-3. The early radial engine did not compare favorably with existing liquid-cooled engines.

Consequently, several years pass before we see the first radial-powered fighter in production, the Boeing P-12 of 1929.

In 1926 we get an inkling of things to come when a single Boeing PW-9 is modified by an engine change and the installation of *wing guns*. This aircraft was designated the XP-4. The experiment apparently was no great success, apparently because wing structures were not yet being designed to accommodate machine guns. However, the significance of the XP-4 lies in that it constitutes tangible evidence that dissatisfaction with synchronization existed and that an early effort was made to overcome its principal disadvantage—the reduction of fire power.

In the early phases of the modern era we must frankly admit that contemporary fighters were hardly more than improved versions of World War I aircraft. We must not expect much in the way of radical design changes until we approach the middle thirties. In the field of fighter armament there is a similar lack of change. Until the late thirties the standard fighter armament consisted of two synchronized machine guns. The caliber .50 machine gun seems to have appeared experimentally during World War I or shortly thereafter. There are instances of some fighters being equipped with one caliber .50 and one caliber .30 machine gun rather than two caliber .30 guns which continued to be standard until the approach of World War II.

The appearance of the Boeing P-12 in 1929 ended the sole domination of the fighter field by the liquid-cooled engine. Until the advent of the jet age the radial engine would share the burden of powering our fighter aircraft. The P-12E of 1932 was considered one of the finest biplane fighters of the world. The same year witnessed the culmination of biplane fighter design employing the liquid-cooled engine in the person of the Curtiss P-6E *Hawk*. When preference for the more acrobatic P-12 was shown, Curtiss countered with XP-23 of 1932, a streamlined version of the P-6E. (Note: Both the P-6E and the XP-23 were developed from the YP-22, our first fighter model to exceed 200 miles per hour.) But the XP-23 was too late; the day of the biplane fighter was drawing to a close. This end was inevitable. The top speed of our biplane fighters was barely 200 miles per hour at a time when monoplane racers were going twice as fast.

THE MONOPLANE FIGHTER APPEARS. The trend toward the low-wing monoplane fighter began with two unsuccessful two-place aircraft, the Detroit YP-24 of 1932 and the Consolidated YP-25 of 1933. Both models are of similar appearance, the YP-24 being of wood and the YP-25 being of all-metal construction. The latter displays the initial use of retractable landing gear in a fighter. It led to the successful P-30 of 1934, which was later redesignated PB-2 and used as an attack plane. But it was the Boeing P-26 of 1933 that eliminated the biplane from further consideration as a fighter. The P-26 is considered by many to have been our first truly modern interceptor. It was a one-place, low-wing, radial-powered monoplane whose speed surpassed that of any biplane fighter in the air forces of the world. We should note in passing that the obsolescence of the biplane was hastened by the appearance in 1933 of the world's first modern bomber, the Martin B-10, whose top speed was in excess of 200 miles per hour. The P-29 of 1934 was Boeing's last entry in the fighter/interceptor field. It was a larger and cleaner version of the P-26, featuring a closed cockpit and retractable landing gear. Although not put into quantity production (only three were built), the P-29 set the pattern for the aircraft which would initiate the latter phase of the modern era.

In the year 1937 four all-metal monoplane fighters were introduced: Seversky (later Republic) P-35, Curtiss YP-36, Curtiss XP-37, and Lockheed XP-38. The P-35 and P-36 were powered by the air-cooled radial engine, the P-37 by a single liquid-cooled engine, and the P-38 by twin liquid-cooled engines mounted in extended nacelles that also acted as booms to support the tail surfaces.

The P-35 originated a series of designs—XP-41 of 1938, P-43 of 1940, and P-44 of 1940—that culminated in the P-47, whose first production model was initially delivered in 1942. The P-36, perhaps our first fighter to see action in World War II, revived the Curtiss *Hawk* series and led to the development of the P-40 in 1938. The P-37 saw only limited production but seems to have suggested the P-40. In its original form, the P-40 was a P-36 modified by changing to the liquid-cooled engine.

The trend toward multiple-gun installation and the incorporation of wing gun-mounting in fighter design had already taken shape by 1939. The Bell XP-39 first appeared in this year, followed by the YP-39 of 1940 which was armed with a 37-mm cannon, two caliber .50, and two caliber .30 machine guns in the nose. The cannon was fired through a hollow propeller shaft as was the case in the Spad 12 of World War I.

The P-36C of 1939 was fitted with two caliber .30 machine guns in the wings. Later experimental models of the P-36 were armed with four and eight wing-mounted guns, and the XP-36F of 1940 had two 23-mm cannon in the wings. This indicates that we were paying heed to British and French complaints that our fighter aircraft were underarmed. In 1941 the P-40D came out with four caliber .50 machine guns in the wings, and subsequent versions of the P-40 carried as many as six similarly located guns. The armament of the XP-46 of 1942, an improved P-40, consisted of ten machine guns.

The XP-47 design of 1940, which never got beyond the drawing board stage, was intended as a light interceptor armed with only two guns. The features of this design were combined with those of the P-44 (which also never got beyond the blueprint stage) to form the XP-47B of 1941. The first production model, P-47B of 1942, carried eight caliber .50 machine guns in the wings. In time the P-47 became our heaviest single-engined fighter, grossing about 10½ tons.

The North American XP-51 of 1941 was originally designed for the British, and there are those who claim that it was designed by the British. The first production model, the P-51 of 1942, was equipped with four 20-mm cannon in the wings. However, later versions were fitted with four and six wing-mounted caliber .50 machine guns.

World War II. Only one more fighter model remains to be introduced to complete the roster of the aircraft that were destined to become our mainstay through World War II. In 1942 the twin-engined XP-61 night fighter-interceptor was introduced. In its original form, the P-61 would have been our most heavily armed fighter of the war, being equipped with four fixed 20-mm cannon and a remote-controlled turret mounting four caliber .50 machine guns. However, the turret was not included in a large majority of the production aircraft.

A larger and more powerful version of the P-39, the Bell P-63, was put into production in 1942. However, practically all of these aircraft went to the Russians and Free French as Lend-Lease equipment.

In the interest of mass production, we limited fighter construction to six basic models at a comparatively early date in the war. These were: P-38, P-39, P-40, P-47, P-51, and P-61. Various modifications of these basic models were incorporated when their nature was such that they caused no great effect on quantity production. But at the same time numerous designs were considered and tested throughout the war. Among these we note such "screwball" designs as *canards* (tail-first aircraft) and tailless flying wings. Also, included in the lot were more serious efforts to improve the basic models.

Under the heading of improved P-40 designs, we note the XP-40Q of 1945, the XP-46 of 1942, and the XP-60 of 1942. Lockheed sought to improve the P-38 in designing the XP-49 of 1942 and the XP-58 of 1943. The latter was similar in external appearance to the P-61; its armament consisted of upper and lower remote-controlled turrets mounting two caliber .50 guns each and either one 75-mm or two 20-mm cannon in the nose.

The XP-72 of 1945, an enlarged version of the P-47 and armed with four 37-mm cannon, never went into production. A later version of this aircraft, designed to intercept buzz bombs in Europe, was cancelled at the end of the war.

Two jet fighters were designed and put into production during the war, the Bell XP-59A of 1943 and the Lockheed XP-80 of 1944. The P-59 was used for developmental work and as a trainer for the P-80 program. The P-80 would have been our first jet fighter of World War II had the war continued.

The Jet Age, 1945 to Date

Due mainly to security considerations, not much publicity was given to the development of jet-powered aircraft here and abroad until after the end of World War II in 1945. Here is where we must pick up and develop the story. And, in the process of doing so, we will consider two periods: the subsonic and the supersonic. The subsonic period had definite limits, a beginning and an end; the supersonic had a beginning, but its ultimate end is limited only by man's ingenuity and resourcefulness.

THE SUBSONIC PERIOD. To trace the history of jet propulsion to its source we would have to go back almost 800 years ago when the Chinese invented gun powder and shortly thereafter made the first rocket by filling a heavy paper tube with the explosive. For purely practical reasons we will begin our story in more recent times but farther back in time than most people realize.

The association of aircraft with the type of propulsion that drives rockets through the air began very early in the history of military aviation. In 1922 our National Bureau of Standards, after analyzing a theoretical jet propulsion system at the request of the Army, reported that "propulsion by the reaction of a simple jet cannot compete, in any respect, with airscrew propulsion at such flying speeds as are now in prospect." One cannot imagine a more perfect squelch; it seems to have ended further serious consideration of the matter in this country for almost twenty years.

But in Europe things were different. In 1928 a young British air cadet, Frank Whittle, wrote a paper in which he predicted the application of jet propulsion to aircraft. Eight years later he began experimenting with the construction of a gas turbine engine. At about the same time similar work was begun in Italy, France, and Germany. Early progress in this field was reported in technical publications; but, when progress assumed practical proportions, the entire area became cloaked in military secrecy. This led to the establishment of a Special Committee on Jet Propulsion by our NACA (National Advisory Committee for Aeronautics) in March of 1941. In the following month General Arnold was amazed when he went to England and saw a jet-powered aircraft built and practically ready for its first flight. This was the Gloster E-28/29, powered by a Whittle-I jet engine; it made its first flight on 14 May 1941. A copy of this aircraft was made by Bell Aircraft and General Electric with the assistance of British engineers including Captain Whittle himself. It was designated the Bell XP-59A and made its first flight on 1 October 1942. This, our first successful jet-propelled aircraft, was actually on the drawing board before the "day of infamy" at Pearl Harbor.

Meanwhile, the Germans had been particularly active in the field of "propulsion by reaction." In 1937 they built the Heinkel HE-112 fighter, a monoplane powered by a conventional reciprocating engine and propeller but with an added 2,200-pound-thrust rocket motor in the tail. This aircraft led to the first rocket fighter, the HE-176 of 1938, and later to the successful Messerschmitt ME-163, which was provided with fuel for 12 minutes and was armed with two 30-mm cannon. The Heinkel HE-178 made the world's first jet-powered flight on 27 August 1939.

Returning back home, we find that production of the P-59A and P-59B was begun

in 1944. The P-59 does not appear to have achieved any great degree of success. Despite its twin jet engines, the P-59 failed to outperform our front line propeller-driven fighters to any great extent. Meanwhile, Lockheed had come along with the P-80 which compared much more favorably. In 1944 the P-80 was placed in quantity production, delegating the P-59 to the principal function of serving as a trainer for the P-80 program. Had the war continued, the P-80 would have seen combat service as our first jet fighter. The end of the war brought with it the cancelling of orders totalling over 4,000 P-80's; over 900 of them were completed and delivered to our air force.

Towards the end of World War II, the Messerschmitt ME-262 made its appearance over Germany. Full production of this twin jet-engined interceptor had been delayed about a year by Hitler's insistance that it be provided with bomb-dropping capability. Had this not been the case, there is no telling what effect the ME-262 might have had on our daylight strategic bombing of Germany in the late phases of the war. About 1,000 of these aircraft were in various stages of construction by the end of the war, most of them being captured or destroyed on the ground. About 100 of them were placed in operation, and several were shot down by our fighters of more conventional design. Not generally known is the fact that a Japanese version of the ME-262, the Kikka, made its first flight on the day after our A-bombing of Hiroshima.

Delivery of additional jet fighter models began in the years immediately following the end of World War II. The prototypes of four aircraft destined to become the mainstays of our fighter/interceptor force arrived in each of four consecutive fiscal years: F-84 in 1946, F-86 in 1947, F-89 in 1948, and F-94 in 1949. Initial deliveries of the production models of the Republic F-84 and North American F-86 were made in 1947, Lockheed F-94 in 1949, and Northrop F-89 in 1951.

In the armament of jet fighters we observe the completion of a cycle in the placement of weapons. High speed demands a thin wing section and, as a result, the wings of jet fighters would no longer accommodate guns and cannon. The elimination of the propeller—and with it the need for synchronization—made the nose of the fuselage the most logical location for weapons. The intended armament for the P-59 had been an assortment of guns and cannon, but with the advent of the P-80 the standard armament for one-place fighters became six caliber .50 machine guns.

The F-84E and F-94A of 1950 were our first fighter aircraft to be equipped with electronic fire control systems. These systems fell into two general classes. In day fighters the fire control system consisted of an A-series optical computing sight and a radar set that provided automatic ranging. In two-place all-weather interceptors a more complex fire control system involving radar sighting was used. Fire control systems of one kind or another became standard equipment in virtually all subsequent fighter/interceptor aircraft. A more recent development in this field is the weapons control system.

The trend to air-to-air rocket armament began with the YF-95A of 1950, which was redesignated YF-86D and put into production as the F-86D in 1951. This special interceptor version of the F-86 carried 24-2.75″ rockets in a retractable tray in the forward underside of the fuselage. Initially the F-94 had been armed with four caliber .50 machine guns in the nose, but the F-94C of 1952 carried a total of 48-2.75″ rockets—24 in the nose, and 12 in each of two wing pods. The original armament of the F-89 had been six 20-mm cannon, but in the F-89D of 1953 this was changed to 104-2.75″ rockets—52 in each of two wingtip pods. Later, the F-89H of 1956 began a new trend in armament with its armament of guided missiles.

THE SUPERSONIC PERIOD. With the coming of the supersonic period we usher in a phase of military aviation to which everything

that has happened in the past is but a prelude. The few paragraphs that conclude our brief history outline the mere beginning of an era which may in time take us into a realm hitherto the exclusive property of the writers of science fiction—space flight. That is the direction in which we have been headed since the first flight of the Wright brothers. We have had to learn to crawl before we could walk, and recent developments seem to indicate that our walking days are over. Space flight may not be "just around the corner," but we have reached a point where the matter can be given more practical consideration than at any time in our history.

Returning to the task at hand—and it does not include gazing into a crystal ball and foretelling the future—we find that the remainder of our history must of necessity be rather sketchy in nature. Now for the first time we are dealing with truly contemporary aircraft and come face-to-face with the limitations of military security.

In the late forties we had several experimental fighters that approached the speed of sound which presented a barrier to further progress. Jet engines possessed a capability for supersonic speeds, but two basic problems stood in the way. First, there was the matter of designing and building wings and airframes of sufficient sturdiness to do the job and yet provide satisfactory control at supersonic speeds. Second, there was the matter of flying through the transonic regions in which shock waves could batter an aircraft out of control or even cause it to disintegrate.

Solution of these problems resulted in a revolution in aerodynamics that radically changed the appearance of aircraft. Wings became thinner, shorter, and stubbier, and were given a pronounced degree of sweepback. Fuselages were stretched out and given extended pointed noses. A later development of this trend was the so-called "coke bottle" fuselage, a most radical departure from subsonic concepts of streamlining. To minimize transonic drag, the fuselage is slightly pinched in near its middle and the aft portion is given a somewhat bulbous appearance.

In 1948 the rocket-powered Bell X-1 successfully passed through the transonic region and attained a speed well beyond that of sound. With this and other similar experimental aircraft we found the answers to the problems mentioned previously. As a result, in 1952, procurement was initiated for the prototypes of the first three of our "Century" series of fighter/interceptors—North American YF-100, McDonnell F-101A, and Convair YF-102. The first flight of a YF-100A occurred on 25 May 1953; that of a YF-102, on 24 October 1953. The world speed record for jet-powered aircraft was broken on 29 October 1953, when a YF-100A attained a speed of approximately 755 miles per hour. (Speed of sound: 741 miles per hour.) To date this is the last specific performance figure released for any of the "Century" aircraft.

The F-101A is a long-range escort fighter for SAC. Subsequently, orders were placed for the F-101B, and interceptor version for ADC, and the RF-101A, a long-range photo-reconnaissance aircraft. Late in 1957 several RF-101A's smashed all existing speed records for both one-way and round-trip crossings of our continent.

In 1953 work was begun on several additions to the "Century" series: Republic XF-103, Lockheed XF-104, Republic YF-105A, and North American YF-107. The XF-103 was produced strictly for experimental purposes. The XF-106 of 1954 was a turbo-prop test model that was later redesignated XF-84H. More recently the designation F-106 has been applied to an advanced successor to the F-102.

Armaments of the various aircraft of the "Century" series vary considerably with respect to the weapons they carry and the method used to control them. Specific details are beyond the scope of this manual at this time.

CHAPTER 2

ARMAMENT FIRE CONTROL PROBLEMS

One of the principal factors in the race for air superiority of interest to us is the competition between offensive and defensive aerial weapons. The interceptor is essentially a defensive weapon. It must cope with offensive weapons whose vulnerability is constantly being reduced by greater speed, greater maneuverability, and defensive protection. To keep ahead in the race, the interceptor must be given even greater speed, maneuverability, and striking power.

The speeds now attained by the interceptor and its target—and the ultimate has not yet been reached—demand that the interceptor deliver its fire power with extreme accuracy at hitherto impossible ranges. This demand has resulted in the development of fire control and weapons control systems—mechanical and electronic devices to assist the interceptor pilot.

Among such devices are automatic sighting and computing systems together with radar equipment that senses the range and direction of targets. The development of these devices has shown a definite trend toward the elimination of the human factor in the solution of the fire control problems. However, security limitations prevent a discussion of the extent to which automatic operation has already been accomplished or will be accomplished in systems now in the process of development.

Because of the aforementioned race, the types of interceptors and their armaments are continually changing. For this reason and because of the security classification of almost all contemporary fire control and weapons control systems, it becomes as important to tell *why* a thing is done as to describe *how* it is done. The *why* properly begins with an explanation of the general fire control problems that the interceptor aircraft must deal with. These problems are basically the same regardless of what particular system we may be concerned with.

Stated in the most simple terms, the fire control problems are to score hits on selected targets using various weapons under varying conditions. While the problems can be stated simply, their solutions are an entirely different matter. Let us see why.

MATHEMATICAL FUNDAMENTALS OF THE FIRE CONTROL PROBLEMS

Before going into the elements and factors of the fire control problems, it is felt that a brief review of certain mathematical fundamentals is in order. The intent is not to prepare you to handle the mathematical solution of specific fire control problems. That is the job of the design engineer. The analysis of the elements and factors of the fire control problems will be more readily understood

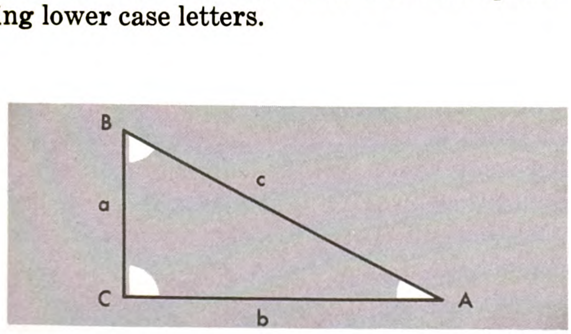
when we get into the later sections of this chapter, but first we must understand certain mathematical concepts such as are involved in elementary trigonometry and vector analysis.

Basic Principles of Trigonometry

Trigonometry is that branch of mathematics that deals with the triangle. As far as trigonometry is concerned, there are only two kinds of triangles—those that are right triangles, and those that are not. The latter are called *oblique triangles*.

Any triangle is made up of six elements: three angles and three sides. When any three of the elements are known—provided that at least one of them is a side—the remaining unknown elements can be calculated by the use of trigonometry. The provision that at least one side be known is important because an infinite number of similar triangles can be formed with a given set of three angles.

Elementary trigonometry deals exclusively with the right triangle, a triangle in which one of the angles is a right angle. It is customary to use a standard procedure in the letter designations of the elements of a right triangle. Capital letters are used to indicate angles, and lower case letters to indicate sides. The *same* letter must be used for an angle and its opposite side as shown in the illustration. The letter C is used to indicate the right angle, and the letter c to indicate the hypotenuse. Which of the two remaining angles is marked A or B is of little consequence, provided that the opposite sides are marked with the corresponding lower case letters.



Basic Right Triangle

Since trigonometry is essentially a practical application of plane geometry, certain simple geometric relationships are carried over. For instance, in any triangle the sum of the three angles must equal 180°. In a right triangle, $A+B=90°$ because $C=90°$. Hence, $A=90°-B$ and $B=90°-A$. Thus, if either of the acute angles of a right triangle is known, the other can be derived by subtracting from 90°.

The other relationship in right triangles is that $c^2 = a^2 + b^2$. This relationship makes possible the calculation of an unknown side of a right triangle when the other two sides are known and both acute angles are unknown. In such cases we would use one of the following formulas, depending on which side is unknown:

$$c=\sqrt{a^2+b^2}$$

$$a=\sqrt{c^2-b^2}$$

$$b=\sqrt{c^2-a^2}$$

TRIGONOMETRIC FUNCTIONS. The ancient Greek geometers knew that there were definite ratios between any two sides of a right triangle and that these ratios varied with the angles. But because of their number system, they could determine the exact ratios for only a few special cases. It was not until comparatively modern times that the advent of decimal fractions made the science of trigonometry possible. We who are concerned with armament may note with interest that much of the early development of trigonometry was due to its practical application to the solution of the fire control problem way back when cannons were first operated beyond point-blank range. In fact, some of the earliest trigonometric works are nothing more than manuals on the direction of artillery fire.

The ratios between the sides of a triangle are called trigonometric functions. Since there are six possible combinations of the three sides taken two at a time, there are six basic functions of any of the angles of a triangle. Their names are: sine, cosine, tangent, cotangent, secant, and cosecant.

Sine. In a right triangle, the sine of an angle is the ratio of the side opposite the angle to the hypotenuse. Hence, $sin\ A = a/c$, and $sin\ B = b/c$. (Note that periods are not used in abbreviating the names of the functions.) The value of the sine is zero when the angle is zero. It increases as the angle increases and reaches a maximum of 1 when the angle is 90°.

Cosine. In a right triangle, the cosine of an angle is the ratio of the adjacent side to the hypotenuse. Hence, $cos\ A = b/c$, and $cos\ B = a/c$. The value of the cosine is 1, its maximum value, when the angle is zero. It diminishes as the angle is increased, and is zero when the angle is 90°.

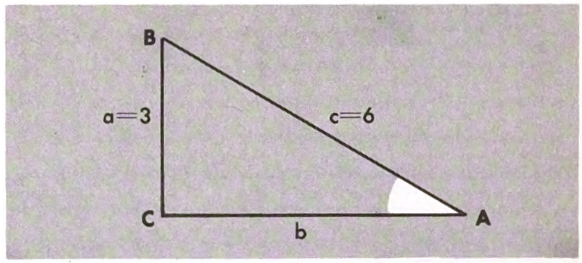
Tangent. In a right triangle, the tangent of an angle is the ratio of the opposite side to the adjacent side. Hence, $tan\ A = a/b$, and $tan\ B = b/a$. The value of the tangent is zero when the angle is zero, and increases as the angle is increased. However, as the angle approaches 90°, the value of the tangent increases at a tremendous rate. As a result, the value of the tangent is infinite at 90°.

Cotangent. In a right triangle, the cotangent of an angle is the ratio of the adjacent side to the opposite side. Hence, $cot\ A = b/a$, and $cot\ B = a/b$. Note that the cotangent is the reciprocal of the tangent.

Secant. In a right triangle, the secant is the ratio of the hypotenuse to the adjacent side. Hence, $sec\ A = c/b$, and $sec\ B = c/a$. Note that the secant is the reciprocal of the cosine.

Cosecant. In a right triangle, the cosecant of an angle is the ratio of the hypotenuse to the opposite side. Hence, $csc\ A = c/a$, and $csc\ B = c/b$. Note that the cosecant is the reciprocal of the sine.

USE OF TRIGONOMETRIC FUNCTIONS. To compute the value of an unknown element, select a function which expresses the relationship between the unknown element and two known elements. For example, a right triangle in which $a = 3$ and $c = 6$ is shown. We would like to know the value of A.



There are two functions that express the relationship between A, a, and c: $sin\ A = a/c$, and $csc\ A = c/a$. Because most common trigonometric tables give only the values of the sine, cosine, and tangent, it is usually advisable to select one of these three functions. Substituting our known values in the sine formula, we get

$$sin\ A = a/c = 3/6 = 1/2 = 0.5000$$

Consulting the tables we find that A, the angle whose sine is 0.5000, is 30°.

If we wish to solve the complete triangle —that is calculate all the unknown elements —the procedure would be as follows:

To solve for B, use the formula $B = 90° - A$. Hence, $B = 90° - 30° = 60°$.

To solve for b, our choice lies between two formulas: $cos\ A = b/c$; or $tan\ A = a/b$. Solving each for b, we get

$$b = c\ cos\ A;\ \text{or}\ b = a/tan\ A$$

Either of the above equations may be used to get the right answer for b. If we substitute values, we find that

$$b = 6 \times 0.8660,\ \text{or}\ b = 3/0.5774$$

The choice of equations is largely a matter of whether you would rather perform the multiplication or the division. Either process will give the same answer: *5.196*.

THE LAW OF SINES. In trigonometry the Law of Sines expresses the relationship between the sides of a triangle and the sines of the angles. The Law of Sines states that the sides of any triangle are proportional to the sines of the opposite angles. This may be

expressed mathematically in the following two ways:

$$a/\sin A = b/\sin B = c/\sin C;$$

and

$$\sin A/a = \sin B/b = \sin C/c.$$

The Law of Sines is especially valuable in solving triangles in which the three known elements are: (1) two sides and an opposite angle; and (2) two angles and any side. Note that the Law of Sines is applicable to *any* triangle.

To use the Law of Sines, it is necessary only to select a portion of the given formulas that contains the unknown and the three known elements, and then to solve for the unknown. For instance, take an example in which B, C, and c are known. To solve for b, begin with the known relationship

$$\frac{b}{\sin B} = \frac{c}{\sin C}$$

Solving for b, we get

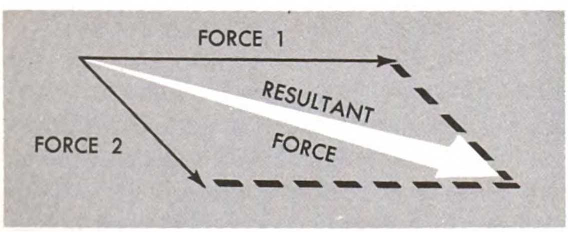
$$b = \frac{c \sin B}{\sin C}$$

Angle A can be calculated by adding B and C, and then subtracting the total from 180°. Once b and A have been calculated, the remaining side can be determined by substituting known values in the formula

$$a = \frac{b \sin A}{\sin B}$$

Vector Analysis

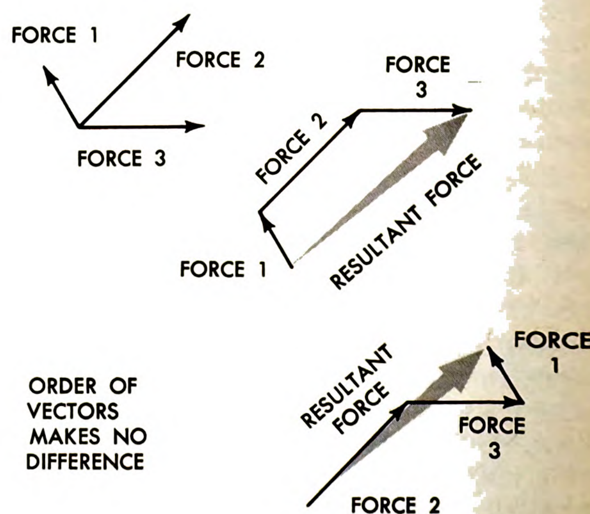
In our discussion of the fire control problems we often encounter situations where an object is acted upon by two or more forces. We will more readily understand what happens to the object if we know some of the fundamentals of vector analysis.



A vector is graphically portrayed as an arrow whose length indicates the amplitude or magnitude of the force and whose direction indicates the direction in which the force is exerted or applied. By using a simple graphic procedure we can resolve any number of vectors into a single resultant.

When only two forces are involved, the parallelogram method is used. First, let a point represent the object. Next, from this point draw two arrows. Make sure that the length of the arrows is proportional to the magnitude of the forces and that their direction is an accurate representation of the direction of the forces. Now, at the tip of each arrow, draw a line parallel to the other arrow, forming a parallelogram as shown in the figure. The direction and amplitude of the resultant force is indicated by the direction and length of the diagonal drawn from the point.

When three or more forces are involved, their vectors are laid out end to end as seen in the related illustration. The resultant vector is drawn from the tail of the first arrow to the tip of the last arrow. The sequence in which the vectors are arranged has no effect on the determination of the resultant.



THE GUNFIRE PROBLEM

Deflection Shooting

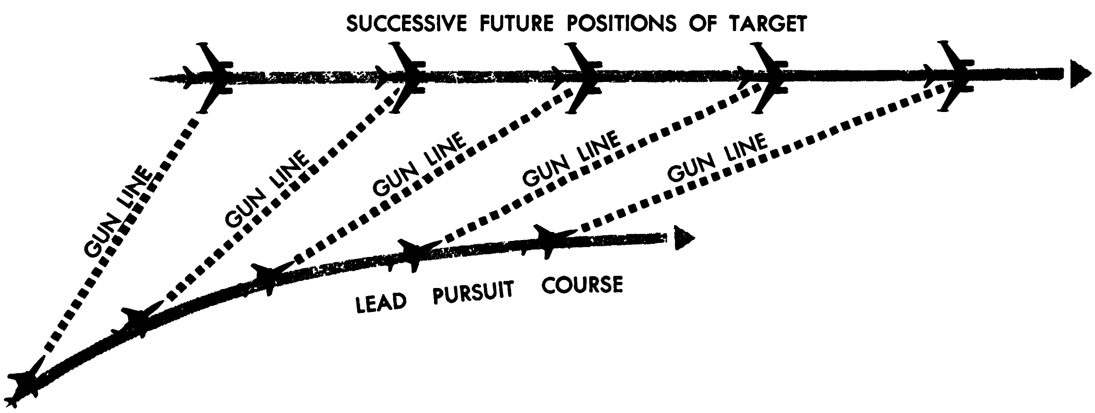
Since the armament of interceptors is fixed to the aircraft and since fire is directed by aiming the aircraft itself, air-to-air firing is usually most accurate when conducted during a direct stern or head-on approach. This fact holds true even though the target presents the minimum area for hits in either of these approaches. However, certain more practical considerations make these approaches undesirable. The high speed and defensive equipment of modern bombers make some type of side approach necessary. And the side approach, in turn, calls for the use of deflection shooting—that is the aiming of the weapons toward a point in space where the target will be at a given instant in the future rather than where the target is right now.

There are two basic types of deflection shooting: the lead pursuit course for aircraft, and the collision course for projectiles. It is essential that we understand the difference between the two because computing systems are usually designed to handle one or the other, but not both.

LEAD PURSUIT COURSE. The lead pursuit course is a spiral approach from the side and usually from the rear of the target as shown in the illustration. This course makes possible the scoring of hits on the target at any time during the approach. Note that the weapons are not aimed directly toward the target, but to a point ahead of the target toward which both the projectiles and the target travel to a collision.

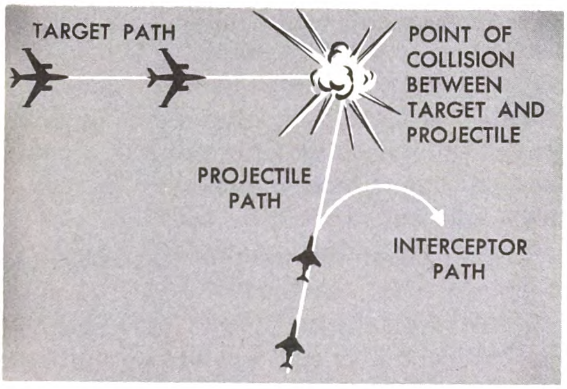
The advantage of being continuously in position to score hits on the target is overcome by two possible disadvantages. First, when the lead pursuit course is conducted from ahead of the target, the rate of closure may be so great that the computing system may not have sufficient time to work out the proper corrections. No matter how much we may speed up the operation of a computing system, the time element can not be completely eliminated. It takes time to analyze fire control problems, and it takes time to derive a solution.

Second, when the lead pursuit course is conducted from astern, the interceptor must be able to overtake the target. If the speed of the interceptor is not appreciably greater than that of the target, there is little prospect for overtaking. In this case the lead pursuit course might be merely a futile gesture. The fact that there is a distinct possibility that attacking bombers can equal or exceed the speed of interceptors explains why there is a trend away from the lead pursuit course and toward the projectile collision course in the design of computing systems.



Lead Pursuit Course

PROJECTILE COLLISION COURSE. The projectile collision course is one in which the interceptor's flight path is caused to intersect that of the target at some fixed point ahead of the target. This course is shown in the accompanying illustration. The point of intersection is roughly the same as the point at which projectiles fired from the interceptor will collide with the target.



Projectile Collision Course

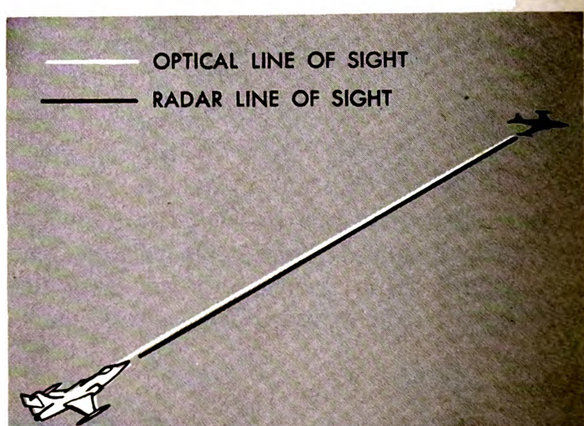
Advantages of the projectile collision course are: (1) that it permits a straight line approach from abeam the target rather than a spiral approach that is difficult for the pilot to maintain with precision; and (2) that, as mentioned previously, it reduces the target speed factor.

On the other side of the ledger lies the obvious fact that this approach provides only a single point from which projectiles can be fired toward collision with the target. Hence, the projectile collision course demands a higher degree of accuracy in the fire control or weapons control system. The target must be blasted out of the sky with the first attack —there may be no opportunity for a second pass.

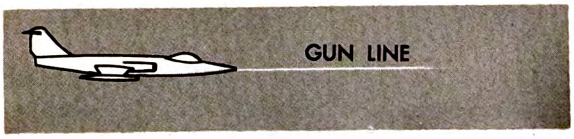
Elements of the Problem

The fire control problem with which we are now concerned is to determine the course of the interceptor, be it aircraft pursuit or projectile collision, that will answer the requirements of the target and interceptor courses and speeds, and of the type of projectiles being fired. The various elements of the problem are discussed in the following paragraphs.

LINE OF SIGHT. From a practical viewpoint, the *line of sight* is the direction in which the interceptor pilot looks or the bearing of the radar antenna when the target is tracked. Hence, the line of sight is an imaginary straight line drawn between the interceptor and the target as shown in the illustration. The importance of the line of sight lies in the fact that the successive positions of this line reveal the manner in which the target is moving with respect to the interceptor. This information is required in the computation of the future position of the target.

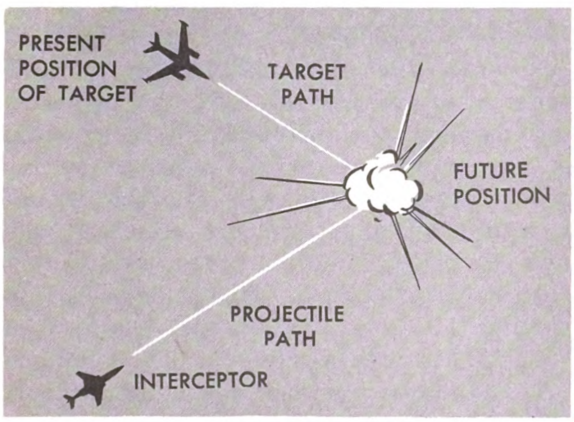


GUN LINE. To simplify our consideration of the fire control problems, it is customary to assume that the computing system is directing the fire of a single gun. The direction in which this single gun is aimed at any given instant is called the *gun line.* Usually the gun line is parallel to the longitudinal axis of the aircraft as shown in the related illustration. It may represent the mean or average of all gun or rocket launcher axes in a multiple armament installation, or it may be a purely arbitrary line that is lo-



cated by installing special fixtures at specific locations on or about the fuselage.

Regardless of how the gun line is determined, any differences between it and actual gun lines are compensated for in the harmonization procedure which will be discussed in a later chapter. In some cases, the gun line is called the *controlled line* or the *fuselage reference line.*



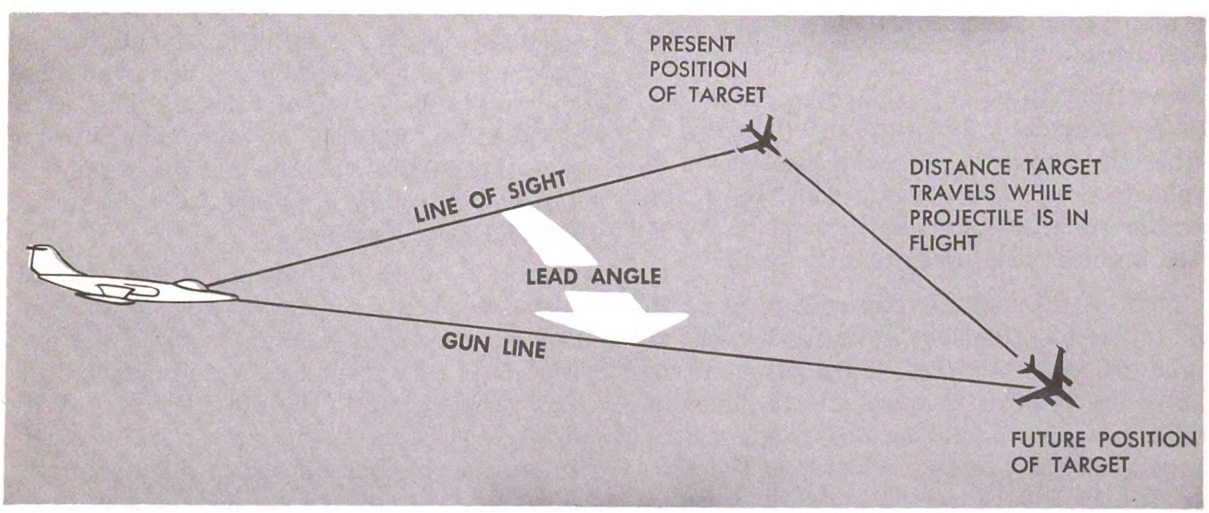
Future Position

FUTURE POSITION. The term *future position* is used only to denote a point in the target flight path where projectiles will collide with the target. As shown in the figure, the future position is always ahead of the target.

The fire control problem would be greatly simplified if the gun line needed only to be aimed at the future position. However, because of the various factors acting on a projectile in flight, the gun line cannot be brought directly to bear on the future position.

LEAD ANGLE. No matter how fast a projectile travels, it requires measurable time to reach a target. During the time that the projectile is in flight, the target moves forward along its flight path. Hence, it is necessary to "lead" the target—that is, aim the gun line ahead of the sight line by enough of an angle that the projectile meets the target at some future point. This angle is called the *lead angle,* and is shown in the captioned illustration. Note that a lead angle can exist with respect to both the horizontal and vertical axes of the interceptor.

ANGULAR VELOCITY. The term *angular velocity* is usually applied only to the angular velocity of the line of sight. It is indicative



Lead Angle

of relative motion of the target with respect to the interceptor and is the rate at which the direction of the line of sight changes as the target is tracked.

During the tracking operation any change in the position of the target relative to the interceptor causes the line of sight to sweep through an angle. The greater the change in relative position, the greater the rate of angular sweep. A sensing device such as a rate gyro may be used to measure the sweep rate and to provide a corresponding output. This output may then be used by the computing system to derive the future position.

In a few words, the computing system derives the future position by extending the observed angular velocity of the line of sight into the future. A high degree of accuracy in predicting the future position demands information relative to possible future changes in the angular velocity. But, for obvious reasons, no tracking system can detect changes *before they occur*. Hence, computing systems are usually designed to anticipate certain future changes in angular velocity. This is generally done by building into the computing system the assumption that the target will continue along its present course and that the interceptor will maintain a particular type of approach. Regardless of what conditions are assumed, the future changes in the observed angular velocity are predetermined and can be compensated for in the computing system.

The limitations of this procedure are easily recognized. Inaccuracy will result if either the target or the interceptor deviates from its prescribed course regardless of the reason. As a result, guided missiles have been applied to interceptor armament.

TIME OF FLIGHT. To know how far into the future to extend the sweep of the line of sight, *the time of flight* of the projectile to the future position must be known. Time of flight depends mainly on the range to the target and the average velocity of the projectile over this range. It can be determined accurately if the muzzle velocity of the projectile, the air resistance, and the range are known.

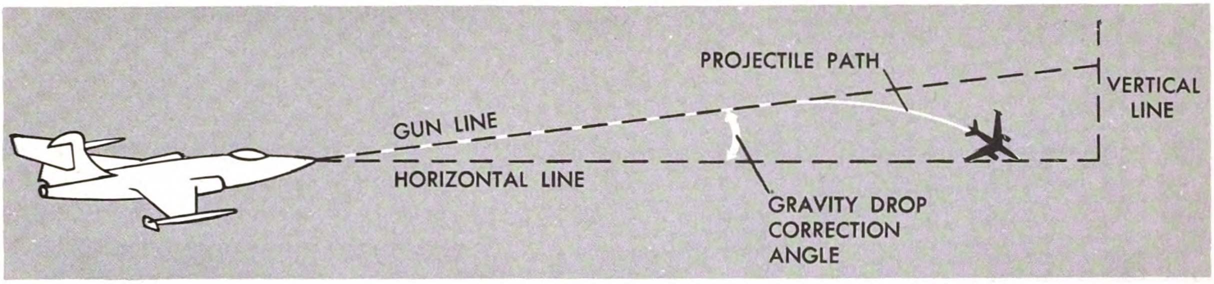
The muzzle velocity of a projectile is known and can be designed into the computing system. (This explains why armament systems are designed for specific weapons.) However, both range and air resistance can vary so much that they must be either measured or estimated, and the computing system must be supplied with corresponding data.

Air resistance, which determines the rate at which the projectile is slowed down as it travels toward the target, is largely a matter of altitude and airspeed. Range is the distance to the target. Hence, the time of flight increases when either the range or the air resistance is increased.

Devices which measure range must necessarily measure the range to the target at the moment the projectile is fired—that is, present range. While some computing systems base the lead angle on present range measurement alone, a more accurate solution is effected when the range to the future position is calculated. This is done by employing a device which senses the rate at which the range is changing. The combination of known present range and the rate at which the range is changing determines what the range will be to the future position of the target.

GRAVITY DROP CORRECTION. The pull of gravity causes a projectile to fall toward the center of the earth from the instant that the projectile is fired or released. This effect is called *gravity drop,* and any change in the direction of the gun line to correct for it is called the *gravity drop correction.* Both are shown in the diagram.

Gravity drop occurs only in the vertical plane—that is, a plane vertical to the surface of the earth passing through the gun line. It is important that we remember this fact because, while the direction of gravity drop never changes with respect to the earth, it can change with respect to the aircraft itself. For instance, the vertical plane of gravity drop is perpendicular to a plane passing

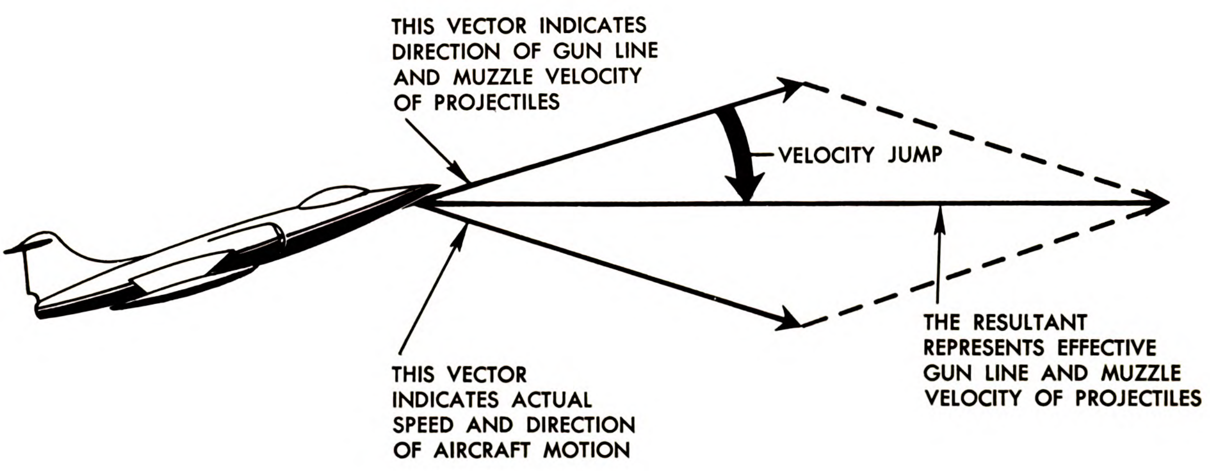


Gravity Drop Correction

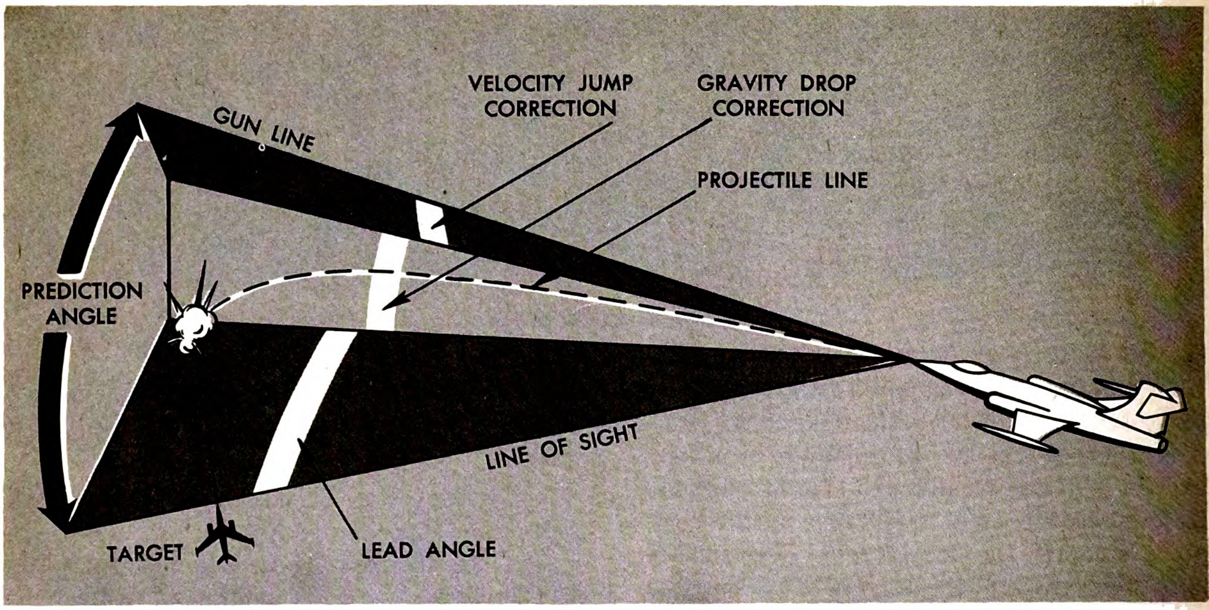
through the wings of an aircraft in level flight. However, the vertical plane and the plane through the wings coincide when the aircraft is in a vertical bank. Since a computing system is usually referenced to the axes of the aircraft, the system must be provided with the capability of transferring the gravity drop with respect to the aircraft axes as the wings are tilted or banked.

Note that the nature of gravity drop is such that the projectile trajectory has a continuously increasing rate of downward curvature. There are two reasons for this. First, the downward pull of gravity causes the bullet to fall faster and faster with the passing of each unit of time. Second, air resistance continuously decreases the speed of the projectile along the gun line. Combining the two effects, we see that during each succeeding time interval the projectile moves a lesser distance away from the gun and a greater distance toward the earth.

VELOCITY JUMP CORRECTION. When the gun line coincides with the direction in which the aircraft is moving, the projectile will initially travel along the gun line with a speed equal to the sum of the muzzle velocity and the aircraft velocity. But often when an aircraft is maneuvering, the gun line and the aircraft direction do not coincide. In such cases the projectile takes an intermediate path, called the effective gun line, that lies between the gun line and the aircraft flight path. The angle between the gun line and the effective gun line, as shown in the diagram, is called *velocity jump*. (It is also sometimes referred to as *trajectory shift.*) Correction for velocity jump is made



Velocity Jump



Prediction Angle

by displacing the gun line through an equal angle in the opposite direction.

Velocity jump lies in the plane of symmetry, which is a plane passing through the center of the aircraft and dividing it into equal halves, right and left. Hence, velocity jump lies in the vertical plane only when the aircraft is in level flight.

PREDICTION ANGLE. The angle between the line of sight and the gun line when the latter is properly aimed to score hits is the *prediction angle*. As shown in the illustration, the prediction angle is a compound angle because it has both horizontal and vertical components. For the sake of simplification, only a horizontal lead angle is shown. It has already been pointed out that both horizontal and vertical lead angles are possible at the same time.

In the illustration, the gun line is offset to the right from the line of sight by the amount of lead and to a lesser degree by the velocity jump. At the same time the gun line is elevated by the gravity drop and velocity jump corrections. The projectile, when fired, takes off from the gun line by the amount of velocity jump. Then, as the projectile travels away from the interceptor, it drops due to the pull of gravity to the target flight path and collides with the target at the predicted future position.

Other factors enter the determination of projectile trajectories, but they are of minor importance compared with those already discussed. The errors they cause are so small that they are well within the tolerance limits of the fire control problems and need no further comment.

THE ROCKETFIRE PROBLEM

Since the elements of the rocketfire problem are essentially the same as those of the gunfire problem, we will discuss only the differences that the rocketfire problem presents.

The Rocket Trajectory

The trajectory of a rocket launched from an interceptor may differ greatly from that of a projectile fired from a gun or cannon on the same aircraft. The gun line has the determining effect on the trajectory of a bullet or cannon shell, but the launcher line

has comparatively little effect on the trajectory of a rocket. Regardless of the direction of the launcher line, the fins of the rocket head the rocket very quickly after launching into the relative wind; that is, along the flight path of the interceptor.

At the moment of launching, the speed of a rocket is largely that of the interceptor. (Compare this with the speed of a bullet or shell whose initial speed is the muzzle velocity plus the interceptor velocity.) However, once the rocket has been launched, its propellant quickly drives it to a high velocity. Meanwhile, gravity pulls the rocket down toward the earth. The pull of gravity acts to depress the rocket line of flight because of the tendency of the rocket to head into the effective relative wind. The effective relative wind is the resultant of two vectors, one of which represents the interceptor flight path and the other of which represents the downward pull of gravity. The result is that the propellant drives the rocket toward the earth faster than would gravity acting alone, and an increasing gravity drop is experienced.

After the propellant has burned out, the rocket assumes a freefall path whose characteristics are determined largely by the velocity at the time of burnout.

The Rocketfire Prediction Angle

Because of the increased gravity drop and the difference in time of flight of rockets over a given range, a computing system designed to accommodate both guns or cannon and rockets would be rather cumbersome. Obviously a different prediction angle would have to be worked out for each different type of weapon.

To date the only fire control systems designed for use with either guns or rockets have considered rockets only with respect to air-to-ground operation. The computers in such systems, when operated in the rocketfire mode, merely modify the gunfire prediction angle. The gravity drop correction is increased according to the type of rocket and the dive angle used. Range is eliminated as a variable factor, and the lead angle is based on the time of flight over the optimum range of the particular type of rocket.

Accuracy in air-to-air rocketfire demands a computer designed specifically for use with rockets, and which handles as variables all of the major factors that affect the rocket trajectory.

Rocketfire Courses

As in the case of gunfire, computers designed for rocketfire operation direct the rockets toward the target from either a lead pursuit course or a rocket collision course. Here, too, the trend is away from the lead pursuit course toward the rocket collision course.

PURSUIT COURSES

When studying pursuit courses, one quickly becomes aware of the fact that the mathematical description of the flight path of an attacking interceptor involves a highly complex equation. For that reason, pursuit courses have been broken down into four types of varying degrees of complexity, depending on how deeply into the subject one wishes to go—the pure pursuit course, the pure lead pursuit course, the aerodynamic pure pursuit course, and the aerodynamic lead pursuit course.

Pure Pursuit Course

The pure pursuit course is the most simple description of the flight path of an attacking interceptor on a curve of pursuit. In the pure pursuit course, the gun line is always pointed directly at the target and a condition of zero angle of attack is assumed. Hence, lead and g loads on the interceptor need not be considered. In this course the only determining

factor is the rate of closure by the interceptor on the target aircraft, which is assumed to be traveling in a straight line. As the range decreases, the flight path of the interceptor tends to become more straight and to approach alinement with that of the target.

Pure Lead Pursuit Course

The pure lead pursuit course is a somewhat more realistic approach. In it, the gun line of the attacking interceptor is offset by an amount determined by the rate of target motion. In other words, the guns are aimed ahead of the target to compensate for target velocity. The pure lead pursuit course is merely the previous course with *lead* added to it.

Aerodynamic Pure Pursuit Course

The aerodynamic pure pursuit course is the curved flight path as actually determined by all aerodynamic considerations. In both the previously described courses, the projectile trajectory is assumed to be directed into the relative wind, that is, the gun line and the direction of interceptor motion coincide. This is definitely not the case when an interceptor is traveling along a curved path. During a coordinated turn an aircraft not only rotates *about* its own axes, but tends to move *along* its axes. As a result, there can be a considerable difference between the direction of the gun line and the direction in which the aircraft is actually traveling when the flight path is curved; hence, the actual projectile trajectory is established not by the gun line but by the effective gun line. The effective gun line is the resultant of vectors representing the gun line and the actual motion of the aircraft during the turn. Therefore, such considerations as lift, trajectory shift, gravity drop, *g* loading, and air density enter into the description.

Aerodynamic Lead Pursuit Course

The aerodynamic lead pursuit course is the most realistic approach and, unless otherwise specified, is the one we refer to when we say lead pursuit course throughout our discussion of the fire control problems. In this type of course, the gun line leads the target by the required amount to score hits and the flight path of the interceptor is actually determined by all aerodynamic considerations. The complexity of the problem of deriving the pursuit course equation makes necessary the logical progression from the simple pursuit course to the aerodynamic lead pursuit course.

LEAD COLLISION COURSE

Because it involves a straight rather than a curved approach to the target, the analysis of the lead collision course presents fewer complexities than the lead pursuit course. To simplify the problem even more, let us consider the lead angle alone.

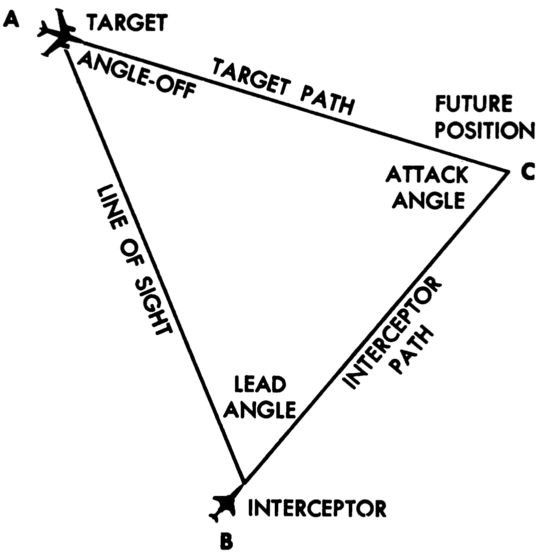
The Lead Angle

The primary point to remember is that the lead collision course is not one in which the target and interceptor flight paths merely intersect at some point. The basis for the establishment of the lead angle in a collision course is one—and only one—definite point at which the two flight paths would result in collision if extended. To analyze this course, let us consider the simple triangle as here illustrated.

In this triangle, *A* is the present position of the target, *B* is the position of the interceptor, and *C* is the point of collision. Then *AB* is the line of sight, *AC* is the target flight path, and *BC* the interceptor course. As far as angles are concerned, *A* is the angle-off, *B* the lead angle, and *C* the attack angle.

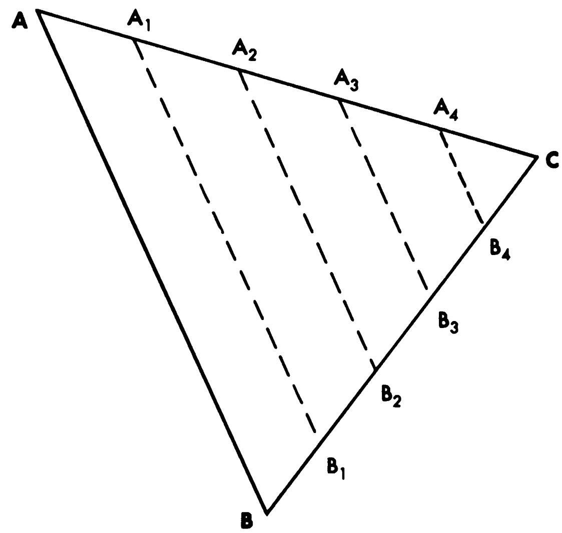
AB, because it is the line of sight, is defined with respect to direction and distance. *AC* and *BC* respectively are the distances the target and the interceptor must travel to the point of collision. For a collision to occur, the target must move distance *AC* and the

interceptor BC in exactly the same time interval. Hence, AC is proportional to the speed of the target, and BC is proportional to the speed of the interceptor.



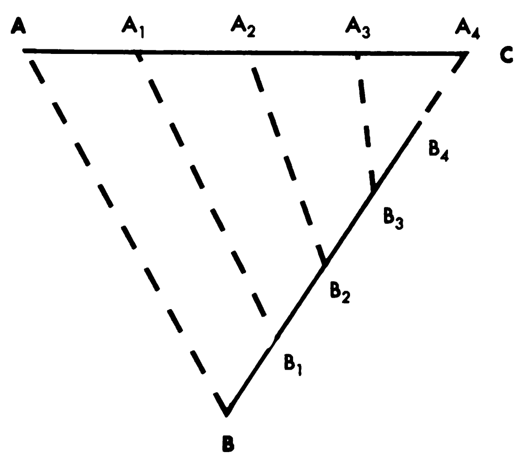
Lead Collision Course Triangle

If both target and interceptor maintain their respective speeds and courses, each approaches the point of collision at a constant rate of change in distance. For example, the target will move halfway toward the point of collision in exactly the same time as the interceptor will move halfway to the same point, regardless of the differences in distances and speeds. Making use of a bit of elementary plane geometry, we observe that at succeeding intervals of time the line of sight together with the remaining distances to the point of intersection form a series of similar triangles as shown in the related illustration. Similar triangles are those in which the corresponding sides are proportional. If the corresponding sides are proportional, the corresponding angles are equal. Hence, $A = A_1 = A_2 = A_3 = A_4$ and $B = B_1 = B_2 = B_3 = B_4$. At the moment we are particularly interested in B, the lead angle. It is the angle between the line of sight and the interceptor course. We can conclude that a collision course can be established merely by heading the interceptor so that the direction of the line of sight to the target *does not change.*



Characteristic of Correct Lead Collision Course

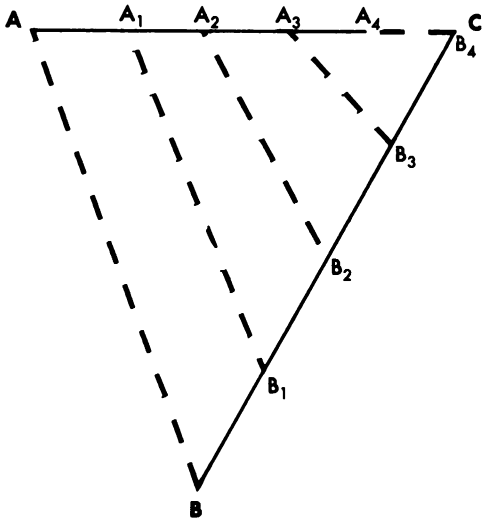
The two simple sketches which follow, show us what happens when a collision course is not established. In the first illustration, the target flight path is divided into



Line of Sight Sweeps Toward Collision Course in Incorrect Lead Collision Course

four equal parts and the interceptor flight path into five equal parts. In this case the target reaches the point of intersection one time unit ahead of the interceptor. After connecting the corresponding points on each flight path, we see that the line of sight sweeps *toward* the point of intersection.

In the second illustration, the target flight path is divided into five equal parts and the interceptor flight path into four equal parts. Now the situation is such that the interceptor reaches the intersection one time unit ahead of the target. Connecting the corresponding points, we see that the line of sight sweeps *away from* the point of intersection.



Line of Sight Sweeps Away From Collision Course in Incorrect Lead Collision Course

Other Factors in the Lead Collision Course

The interceptor will discharge its weapons toward the target at some point along its course prior to the point of collision. These projectiles will travel the distance to the point of collision in less time than it would take the interceptor itself. As a result, the projectiles could reach the point of collision ahead of the target. Hence, the computing system must compensate for the speed of the projectiles. In addition, the computer must correct for gravity drop and any other factors that affect the trajectory of the projectile as is the case in the lead pursuit curve.

The 90° Beam Attack

The 90° beam attack is a special case of the lead collision course in which the attack angle—the angle between the target course and the interceptor course—is exactly or approximately 90°. The advantages offered by this type of attack are such that we may see a more widespread application of it in the not-too-distant future. Let us see what these advantages are.

First, the margin by which the interceptor will physically miss the target after firing is quite small in any lead collision course. This is the inevitable result of there being little difference between a collision course for projectiles and a collision course for aircraft. This difference is maximum in the 90° beam attack.

Second, the 90° beam approach tends to keep the interceptor clear of sectors covered by any nose or tail radar-controlled fire-control system the target aircraft may be carrying. Further, the use of this attack angle may greatly enhance the element of surprise when the target aircraft is equipped with advanced warning or fire control radar.

Third, the 90° beam attack makes it possible to keep a target of any speed within the azimuth limits of the radar search pattern. Radar contact cannot be expected when the target is at extreme angles with respect to the nose of the interceptor. This is especially true when the interceptor approaches the target from the rear and at approximately the same speed.

Fourth, the 90° beam approach reduces the fire control problem to its most simple form. When the attack angle is 90°, the sum of the angle-off and the lead angle equals 90°. Further, the angle-off is the angle whose tan-

gent is the ratio between the speeds of the interceptor and the target. Hence, the computing system can derive the basic lead angle merely by establishing the speed ratio.

In the beam approach the attack angle need not be exactly 90°. Practical considerations make a tolerance of $\pm 10°$ allowable.

FACTORS AFFECTING THE AIR-TO-AIR ATTACK PROBLEM

All the factors that affect the air-to-air attack problem are important. If any of them were not, we would not have to bother with it. It is difficult to say that any one factor is more important than any other because the relative importance can vary considerably with individual circumstances. However, all other things being equal, range generally is the dominant factor in the determination of the prediction angle. From a historical point of view, the truth of this statement is evidenced by the fact that the principal weakness of early fire control systems lay in the ranging devices used with them. In fairness it must be pointed out that the blame cannot be laid entirely on the devices alone—the human element was a large factor. Under combat conditions, the manual operation of a ranging device with the required degree of accuracy is often quite difficult. True accuracy in fire control systems had to await the arrival of automatic radar ranging.

Range

Range is a factor in two of the three corrections that compose the prediction angle—lead angle, and gravity drop correction. The lead angle is determined by the time of flight and the angular velocity of the line of sight. The time of flight, in turn, is determined by range and air resistance. Increasing the speed or altitude of an aircraft can diminish the effect of air resistance on the time of flight. But, if we disregard the effect of air resistance, time of flight is always proportional to range.

Gravity drop is largely a matter of time of flight. Disregarding air resistance again for a moment, we see that the rate at which a projectile drops increases proportionally to the square of the time since the projectile was fired. Hence, the distance below the gun line that a projectile drops depends on time of flight. Since the relationship between time of flight and range has already been stated, we can say that gravity drop depends on range.

The difference between present range and future range has already been discussed. In earlier fire control systems, computing systems dealt with future target position in terms of present target range. No great inaccuracy resulted from this procedure when dealing with reciprocating-engine-driven aircraft, using the lead pursuit course. In most cases the difference between present and future range did not exceed 100 feet.

However, in dealing with jet aircraft and lead collision courses, the difference between present and future range can be quite considerable. As mentioned previously, future range can be calculated by taking into consideration present range and range rate (the rate at which range is changing).

Angle-off

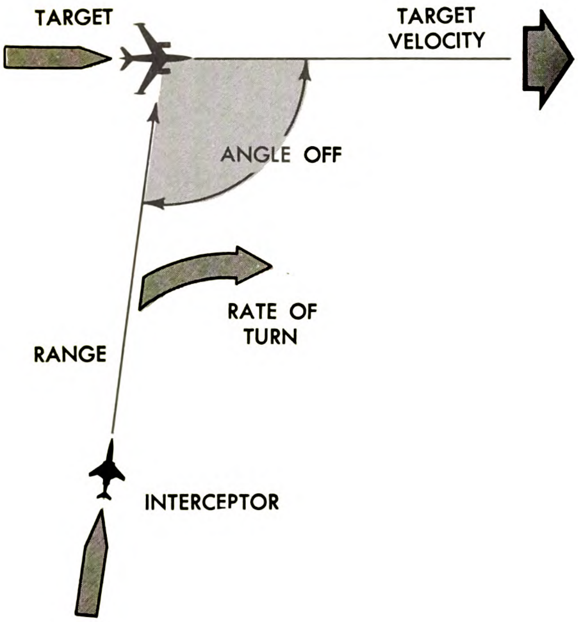
Angle-off, the angle that the line of sight makes with the target flight path, is a principal factor in establishing the rate at which the interceptor must turn during a pursuit course. Let us analyze the rate of turn and then see how it affects the total *g* force acting on the interceptor.

RATE OF TURN. In a pursuit course, the rate of turn is determined by the angle-off, target velocity, and range. The rate of turn, here diagramed, is expressed by the formula

$$\textit{Rate of turn} = \frac{\textit{Sine of angle-off} \times \textit{target velocity}}{\textit{Range}}$$

Taking the factors one at a time, we see that the rate of turn increases when the target

speed increases. The faster the rate at which the target passes across the path of the interceptor at any given moment, the faster the interceptor will have to turn to keep its gun line on or near the target.



Rate-of-Turn Factors

Next, the rate of turn decreases as the range increases. Consider the case of two targets traveling in the same direction and at the same speed but at different ranges from the interceptor. In any given time interval both targets will travel the same distance. But the interceptor turns through a smaller angle to keep its gun line on or near the more distant target.

Finally, since it is proportional to the sine of the angle-off, the rate of turn is maximum when the angle-off is 90°. Of further interest is the fact that the formula demonstrates that the rate of turn is zero during a direct head-on or tail approach. At 0° and 180° the value of the sine is zero, so the rate of turn is zero regardless of what values we may have for target velocity and range.

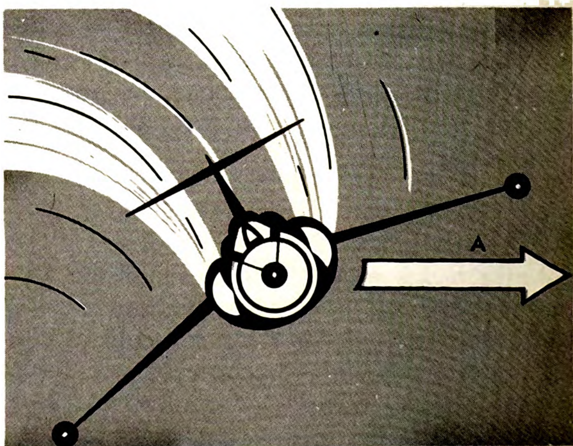
The formula as presented gives the rate of turn in radians per second when both the target velocity and the range are expressed in terms of the same unit distance, such as feet per second and feet respectively. A radian is 57.296°.

RADIAL ACCELERATION. Centrifugal force acts on an aircraft beginning a turn. That is, an acceleration in the horizontal plane is combined with the force of gravity. This acceleration is called radial acceleration and is designated by the letter A in the following diagram. It is expressed mathematically by the formula

$$\text{Radial Acceleration} = \frac{\text{Rate of turn} \times \text{Interceptor Velocity}}{\text{Gravity acceleration}}$$

Note that radial acceleration is directly proportional to the rate of turn, which is proportional to the sine of the angle-off. Therefore, there is a relationship between radial acceleration and angle-off.

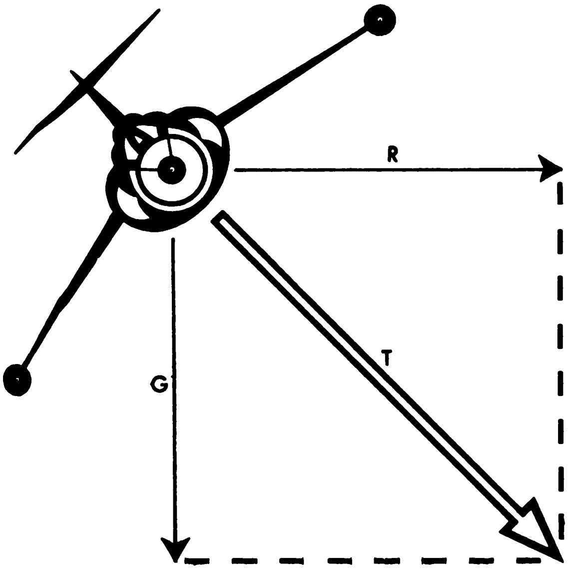
In determining radial acceleration, the interceptor is assumed to be making a level turn. Hence, the value for gravity acceleration is equal to 1 g, or 32.2 feet per second squared.



Radial Acceleration

TOTAL G FORCE. The total g force is the resultant of vectors representing radial acceleration, and the weight of the aircraft is

equal to 1 g or 32.2 when a level turn is made. If the interceptor climbs while making the turn, the effective weight of the aircraft increases. If the interceptor dives while making the turn, the effective weight decreases. In any event the total g force, T, always acts along the vertical axis. Our mathematical development shows that the total g force is related to the angle-off.



Total G Force

Altitude

Altitude is both a direct and indirect factor in the air-to-air attack problem. Its direct effect is the decrease of the time of flight as the altitude is increased, all other conditions remaining unchanged. The decrease in the time of flight is purely a matter of the decrease in air density with an increase in altitude. The less dense air is, the less resistance it offers to a projectile in flight.

An indirect effect of altitude is the decrease in the gravity drop of a projectile as the altitude increases, all other conditions remaining unchanged. It has already been shown that gravity drop is purely a matter of time. Hence, if altitude decreases the time of flight, it cannot help but reduce the gravity drop over a given range.

Perhaps the most simple way to demonstrate the effect of altitude on time of flight and bullet drop is to present a table in which the only variable is altitude. In the following table the interceptor velocity is 450 mph, range is 4,000 feet, and muzzle velocity is 2,900 fps.

Altitude (feet)	Time of Flight (seconds)	Bullet Drop (inches)
0	1.557	379.6
7,000	1.438	340.2
16,000	1.339	308.6
28,000	1.259	283.7
44,000	1.199	265.6

Air temperature is usually associated with altitude. Normally we expect the air temperature to decrease as the altitude is increased. However, because local conditions enter into the picture, the relationship between altitude and temperature is subject to variation. For instance, an interceptor flying at a constant altitude may pass through air masses having different temperatures.

What we are leading up to is the fact that the muzzle velocity of a projectile varies with the temperature of the propellant powder charge. If the temperature of the powder charge is reduced, the muzzle velocity is also reduced. Since the muzzle velocity determines the average velocity over any given range, any change in the muzzle velocity will affect the time of flight.

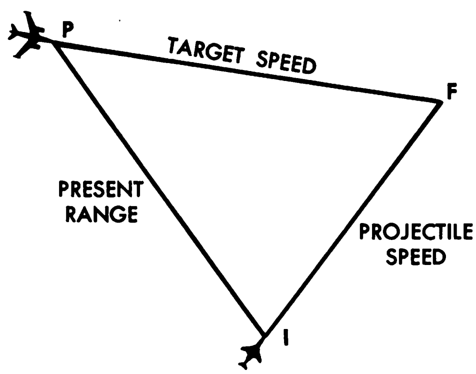
To date no computing system has compensated for the effects of air temperature. Since all-weather interceptors are called upon to operate under varying temperature conditions, it may become necessary to incorporate compensation for temperature variations into computing systems. The alternative seems to be to apply heaters to gun and ammunition spaces in interceptors.

Target Speed

The effect of target speed can be shown in a simple trigonometric relationship. Take

the triangle shown in which the three vertices are the present target position P, future target position F, and interceptor position I. Hence angle P is the angle-off, angle I is the lead angle, side PF represents target speed, and side IF represents projectile speed. Using the Law of Sines, we can derive the formula

$$\sin I \text{ (lead angle)} = \frac{PF \text{ (target speed)} \times \sin P \text{ (angle-off)}}{IF \text{ (projectile speed)}}$$



Effect of Target Speed

Hence, the lead angle is directly proportional to target speed multiplied by the sine of the angle-off. Thus, the effect of target speed is maximum when the angle-off is 90°. It is minimum (zero) when the angle-off is 0° or 180°. As we might expect, the lead angle decreases as the average projectile speed is increased.

From a purely practical point of view, a computing system cannot use target speed directly in the generation of the lead angle. As mentioned previously, the lead angle is derived by projecting the rate of change in the direction of the line of sight and the rate of change in range into the future. Both of these rates are determined by the combined effects of target speed, interceptor speed, and angle-off. There is no need for taking these effects individually.

Although there are other complicating factors, the principal effect of interceptor speed is on projectile time of flight and gravity drop. This can be shown by a table in which interceptor true airspeed is the only variable and other factors such as altitude and range are held constant.

True Airspeed (miles per hour)	Time of Flight (seconds)	Bullet Drop (inches)
0	1.688	482.1
150	1.554	410.8
300	1.439	354.3
450	1.339	308.6
600	1.253	271.3

The results shown in the table are what we normally would expect. The average projectile speed is not determined by muzzle velocity alone, but by the sum of muzzle velocity and interceptor speed. As a result, the greater the interceptor speed, the faster the average projectile speed. Hence, increasing the interceptor speed reduces the time of flight and also the gravity drop.

Type of Armament

The effects of the various ballistic factors on various projectiles are so different that it is customary to design computing systems to accommodate specific weapons. The following table shows clearly the difference between several representative projectiles in time of flight and gravity drop over the same range and under approximately the same conditions.

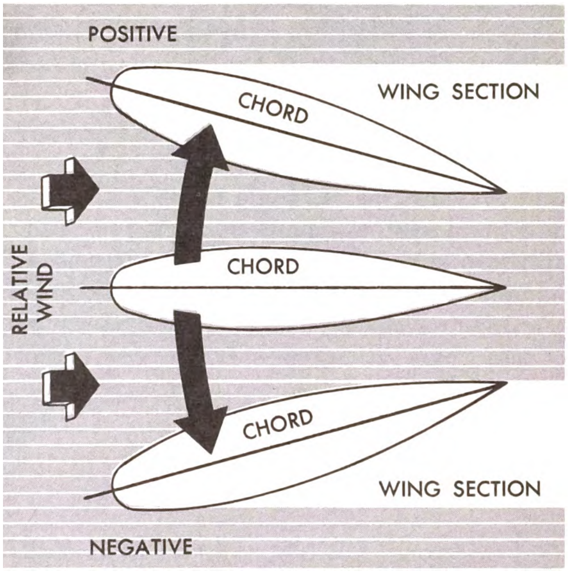
Projectile	Propellant Temp. (Degrees)	Time of Flight (Seconds)	Gravity Drop (Inches)
Cal. .50	--------	1.446	330.2
20 mm.	--------	1.58	386
2.75″ Air Rocket	10°	6.571	2,700
2.75″ Air Rocket	70°	6.419	2,500
2.75″ Air Rocket	130°	6.265	2,300

Note the decided effect of propellant temperature in rockets. This effect is also present in the case of cal. .50 bullets and 20-mm shells, but to a lesser degree. We can expect the effect of temperature to increase as the size of the powder charge increases. In cases where a computing system was used with more than one weapon, the basic computer was designed to accommodate one specific weapon. To this were added other devices which modified the prediction angle computed for that weapon so that the modified prediction angle was accurate for the other weapon under limited conditions. Space and weight requirements normally prevent the equipping of an aircraft with the separate and complete computing systems necessary for operation of more than one type of weapon.

Angle of Attack

The angle of attack should not be confused with the previously discussed attack angle, which is the angle between the target flight path and the interceptor flight path. The angle of attack is the small angle between the chord of the wing and the direction of the relative air through which the interceptor moves. The angle of attack may be either positive or negative. It is positive when the chord is tilted so that the leading edge is above the trailing edge, and negative when the leading edge is below the trailing edge as shown in the related illustration.

The lift provided by a wing varies with the angle of attack and the speed at which it travels through the relative air. The lift increases as the speed increases. The lift also increases as the attack angle increases in the positive direction, reaching a maximum value at some value depending on the airfoil section. When the angle of attack is increased beyond this value the *stalling point* is reached and lift drops off sharply. Changing the angle of attack in the negative direction decreases the lift. At some negative value which depends on the airfoil section, the lift becomes zero.



Positive and Negative Angle of Attack

In level flight the speed and angle of attack must be such that the lift neutralizes the weight of the interceptor. Suppose it is desired to increase the speed and still maintain level flight. Merely to increase the speed will increase the lift of the wing and cause the interceptor to rise as well as go faster. Hence, to maintain level flight the angle of attack must be decreased so that lift is decreased by the same amount that it was increased by the increase in speed.

Now we come to the $64,000 question: How does angle of attack enter as a factor in the fire control problems? First of all, let us analyze the situation. The attitude of the wing is subject to change with aircraft speed. Furthermore, attitude of the entire aircraft is subject to change with speed. Now, take into consideration the fact that the gun is fixed with respect to the fuselage. The only conclusion is that the direction of the gun line and the direction in which the aircraft travels through space are not necessarily the same and that the difference varies with the speed of the interceptor —*even during level flight.* The result is similar in nature to velocity jump in that the

projectile trajectory initially travels along the resultant of vectors representing the gun line and the actual direction of aircraft motion.

Accuracy in the solution of the gunfire problems demands compensation for the effect of variation in angle of attack. Early fire control systems made no such provision. Later computing systems were designed to compensate for variation in angle of attack by providing an on-the-ground adjustment for any one of several fixed interceptor speeds. More recent computing systems provide for automatic inflight compensation for continuous variation in the angle of attack.

Relationship of Fixed Gun Line to the Fuselage Reference Line

The relationship between the fixed gun line and the fuselage reference line is involved in the harmonization process and will be discussed fully in a later chapter dealing with that subject.

CHAPTER 3

COMPONENTS OF COMPUTING SIGHT SYSTEMS

In this chapter we are primarily concerned with optical sighting systems. Despite the definite trend away from such systems, there is a possibility that they may continue in use—to a more and more limited degree, perhaps—for some time to come. We must be careful not to let any consideration of obsolescence prejudice us against optical sighting systems. Many basic devices used in optical systems have been incorporated into the so-called modern systems. One of the basic truths in the field of science is that fundamental principles never become obsolete. The most up-to-date vacuum tube in a modern radar set is a brother of the first crude triode made by DeForest in 1906, and the underlying principle of both was discovered by Edison in 1883.

Our discussion of optical sighting systems will be preceded by a brief review of optics and the transmission of light. This review will be limited to those principles which are pertinent to optical sights, and our approach will be as nonmathematical as possible.

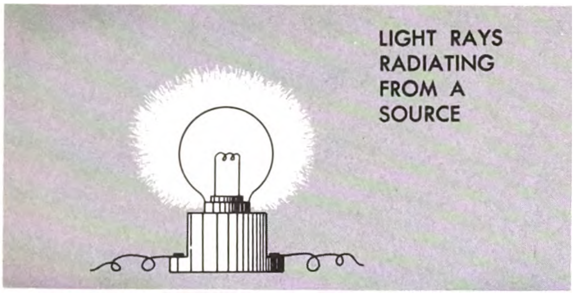
OPTICS INVOLVED IN SIGHTING SYSTEM

Light and Its Transmission

Light is a form of energy that radiates in all directions from its source. We must not be too curious as to the exact nature of light because scientists are in disagreement. Nor can they give us a satisfactory, simple explanation of why light does some of the things it does. For instance, if we assume that light is transmitted as waves through a medium, as is the case with sound, how can we explain the transmission of light through a vacuum? In some respects the behavior of light is quite similar to that of radio waves, but there are sufficient differences to make us exercise caution in drawing analogies between the two.

Since we are not concerned with what happens to light out in the dim distant reaches of outer space, we can safely assume that light travels in straight lines. Hence, we can graphically portray a light source as a point and the light itself as a group of straight lines or *rays* radiating from the point or source. To be truly accurate, we would have to show light as an extremely large number of lines radiating from the source, as shown in the accompanying figure. To simplify our diagrams illustrating the transmission of light, we will use only the minimum number of rays that will indicate the behavior of light.

In a vacuum light travels at the same speed as radio waves—186,000 miles per second. In air of average density the speed is slightly slower, but the difference is so small that the same figure can be used with



no great degree of inaccuracy. However, the most important point for us to remember at this time is that the speed of light varies with the density of the medium through which it passes. The more dense the medium, the slower the speed of light passing through it. For example, the speed of light through water is about three-fourths of that through air. Through glass it varies from about one-half to two-thirds, depending on the composition of the glass.

There are two basic methods of changing the direction of light rays: reflection and refraction.

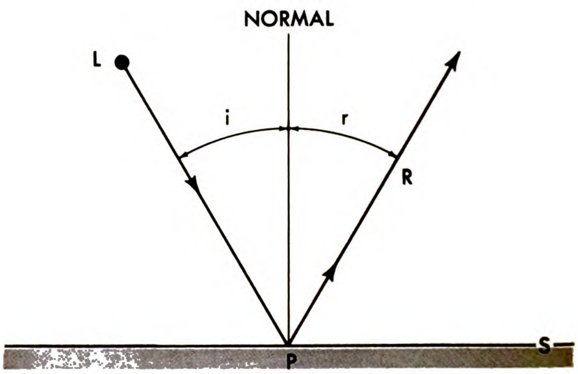
Reflection occurs when light rays "bounce" off a smooth shiny surface. Normally we associate reflection chiefly with mirrors and the like. But we must remember that any surface that is neither rough nor dull is capable of causing reflection. Consider for a moment the fact that, long before the invention of mirrors as we know them today, the Romans created reflective surfaces by smoothing and polishing sections of ordinary plaster walls.

Refraction is the "bending" of light rays when they pass through transparent substances of different densities. Perhaps the most common example of refraction is the interval of twilight between sunset and dark. After the sun has dropped below the horizon, some of the rays of sunlight are bent earthward in passing from the vacuum of outer space into our atmosphere. Were it not for this refraction, darkness would occur immediately after sunset.

Special mention must be made of certain plastics which "conduct" light. For instance, if a light source is placed at one end of a rod made of any one of several plastic materials, practically all the light that enters the rod will be transmitted through the rod and come out of the other end—even though the rod is curved. This type of light transmission is made possible by refraction which bends the light rays back and forth between the outer surfaces of the material. However, some light rays will pass through the outer surfaces of the plastic if not provided with an opaque coating.

Reflection

PLANE SURFACES. The basic law of reflection states that the angle of reflection is equal to the angle of incidence. What this means can easily be shown graphically. In the diagram, *S* represents a smooth plane surface and *L*, a light source. A single ray of light is represented by the solid line running from *L* to *P*. The light line through *P* is a normal (line perpendicular to *S*). The direction of the reflected ray *R* is such that *r*, the angle of reflection, is equal to *i*, the angle of incidence.

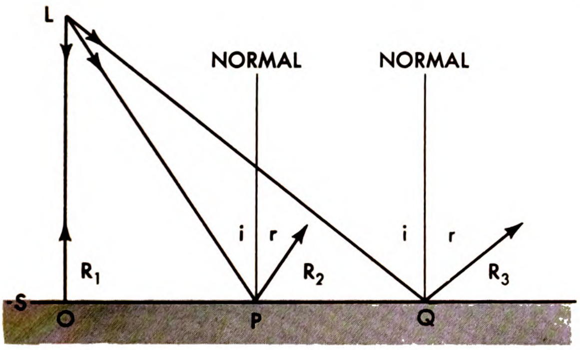


Individual rays of light coming from a single light source do not necessarily meet a plane surface at the same angle of incidence. The next diagram shows three such rays meeting the plane surface at points *O*, *P*, and *Q*, respectively. If the light ray coincides with the normal as at *O*, both the angles of incidence and reflection are zero. In this

case, the light ray is reflected directly back toward the source. The angle of incidence at *Q* is greater than at *P*. However, the angle of reflection is equal to the angle of incidence at each point.

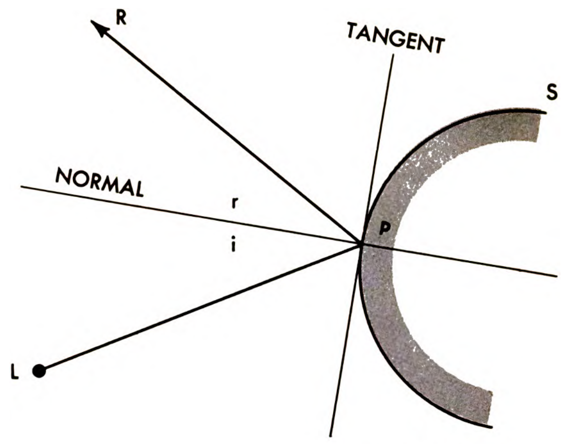
NOTE

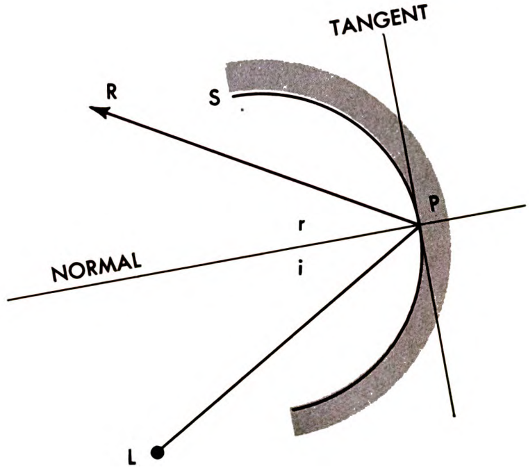
An exception to the above is observed when the sun is considered as a light source. Because of the tremendous size of the sun and its distance from the earth, the sun's rays are considered parallel to each other. Hence, when we deal with rays of sunlight, the angle of incidence is the same at all points of a plane surface.



CURVED SURFACES. The law of reflection also applies to curved surfaces, both concave and convex. However, there is a difference which results from the fact that a line cannot be drawn perpendicular to a curve. Hence, the normal must be drawn perpendicular to a line that is tangent to the curve at the desired point. The diagrams illustrate reflection from both a concave and a convex surface.

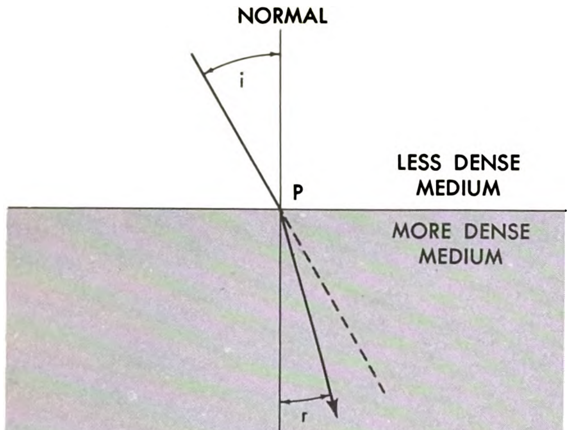
You may find it convenient to know that there is a shortcut method of locating a normal to a curve, provided that the center of the curve is known or can be accurately determined or estimated. A line passing from the center of the curve through the desired point is the normal to the curve at that point.

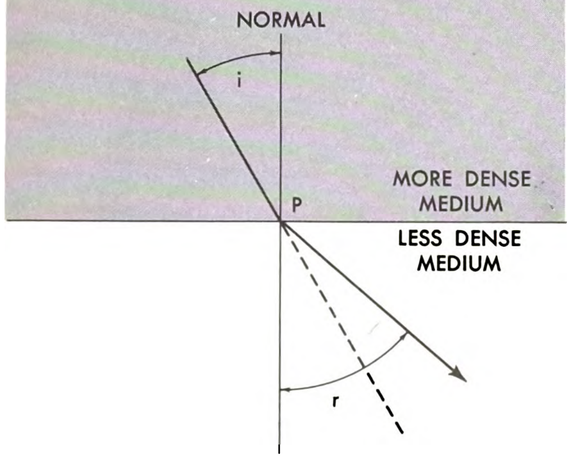




Refraction

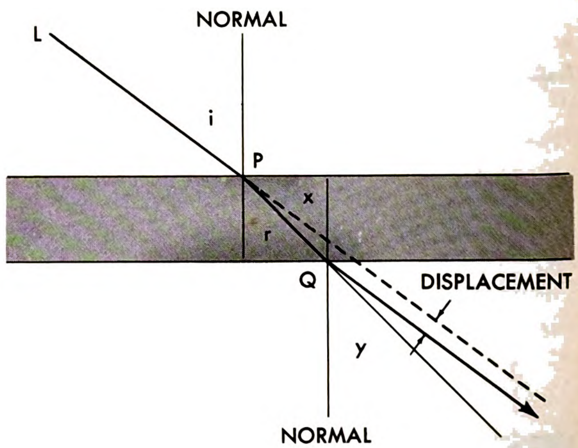
THE LAW OF REFRACTION. Refraction is the change in direction that takes place when light passes obliquely through mediums having different densities. The law of refraction states that light rays are deviated toward the normal when passing from the less dense to the more dense medium; conversely, they will be deviated away from the normal when passing from the more dense to the less dense medium. In other words, the angle of refraction is less than the angle of incidence when light enters a more dense medium; and the angle of refraction is greater than the angle of incidence when light enters a less dense medium. This is shown in the diagrams.





Refractive Index. The relationship between the angle of refraction and the angle of incidence is not quite as simple as the one we saw in the case of reflection. It is such that the ratio between the sines of the two angles is equal to the ratio between the refractive indexes of the two respective mediums. The *refractive index* of a medium is a number that expresses the ratio between the speed of light through a vacuum and its speed through the medium. For instance, the refractive index of water is 1.333, which means the speed of light is 1.333 times greater through a vacuum than it is through water.

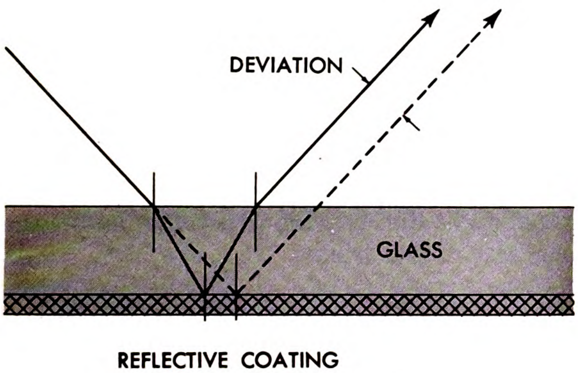
Refraction in a Plate or Slab. When light passes through a parallel-sided transparent plate or slab such as a piece of glass, it is displaced laterally but it is not deviated. Consider the accompanying diagram. A ray from light source *L* meets the upper surface at point *P*. Since it enters a more dense medium, the ray is deviated toward the normal. Hence at point *P*, angle *r* is smaller than angle *i*. The ray emerges from the plate at point *Q* and returns to the original medium. In passing into the less dense medium the ray is deviated away from the normal. Hence, at point *Q* angle *y* is greater than angle *x*.



Since the light passes from one medium to another and then back to the original medium, we can logically expect the deviation at point *Q* to be equal and opposite to that at point *P*. This is the case and, as a result, the angle of refraction *y* is equal to the angle of incidence *i*. Thus, the ray of light travels in the same direction after passing through the plate, but it has been displaced laterally.

Front-Surface Mirrors. An example of refraction that is more pertinent to the subject of optical sights will be discovered if we analyze what takes place in an ordinary glass mirror. In such devices a reflective coating is placed on the back side of a glass plate. Hence, a ray of light must be refracted —even if only slightly—both on entering

and leaving the front surface in addition to being reflected by the back surface. The diagram shows what happens, but it must be admitted that this effect is exaggerated. The solid line shows the path of a light ray, and the dashed line shows what the path of the light ray would be without refraction. Note that the ray is displaced but not deviated.



Refraction in a Rear-Surface Mirror

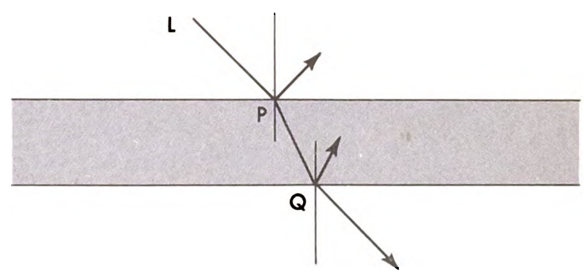
The effect of refraction is so small that we do not notice it when we stand in front of a mirror to shave, comb our hair, or straighten our neckties. But nevertheless the displacement can be detected with instruments and must be taken into consideration in the design of accurate optical systems. As a result, some sighting devices make use of front-surface mirrors. As the name implies, these are mirrors which have the reflective coating applied to the front surface instead of the back surface. Hence, the mirror provides reflection without refraction. When cleaning front-surface mirrors, technical order procedures should be followed carefully to prevent damage or injury to the exposed reflective coating.

Prisms

Prisms are solid figures whose ends are identical and parallel polygons and whose faces are parallelograms. When used for optical purposes, they are made of glass or some other suitable transparent material. Triangular prisms—that is, prisms whose bases are triangles—are most commonly used in optical devices.

REFLECTION BY PRISMS. Normally we associate reflection with mirrors. However, prisms can be used in place of mirrors to change the direction of light by reflection. From a military viewpoint, there are several advantages in doing so. Of these, the two most important are: (1) prisms are much more rugged than mirrors and are, therefore, less subject to damage; and (2) prisms require no reflective coatings as is the case with mirrors.

Total Reflection. An effect known as *total reflection* makes possible the use of prisms in place of mirrors. To prevent confusion, no mention was made in our earlier discussions that refraction is accompanied by reflection. To see what is meant by this, let us consider the diagram.



Refraction Can Be Accompanied by Reflection

A light ray from source L meets the upper surface of a glass plate at point P. Most of the light will be refracted toward the normal as discussed previously. However, part of the light will be reflected by the upper surface as indicated by the short arrow at P. After traveling through the plate, most of the light will be refracted out into the original medium at point Q. But a portion of the light will be reflected back into the glass by the lower surface as shown by the short arrow at point Q.

Under certain conditions, a light ray reaching the boundary between a more dense

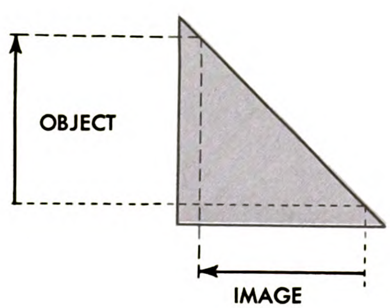
and a less dense medium is not refracted at all—instead, all the light is reflected. This effect is called *total reflection.*

Total reflection occurs when the angle of incidence is greater than a certain value known as the *critical angle.* A close approximation of the critical angle of any medium can be derived by finding the angle whose sine is equal to the reciprocal of the refractive index of the particular material. This expression looks much more simple than it sounds when set up as a formula

$$\sin C = \frac{1}{m}$$

where C is the critical angle and m is the refractive index of the material. Approximate values for the critical angles of the two types of glass most commonly used in optical devices are: flint glass, 30° to 40°; and crown glass, 38° to 42°. The spread in the given values results from the fact that glass is a synthetic material and, therefore, its composition and its characteristics vary.

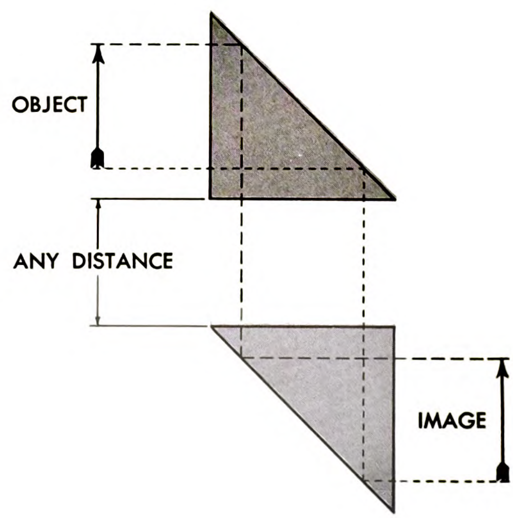
Uses of Prisms as Reflectors. When used as a reflector, a prism may be designed to perform any of the following functions: turn light 90°, turn light 180°, or erect an inverted image. We must be careful to note that these are not all the functions that can be performed by a prism; we are merely limiting ourselves to a discussion of the more common functions and to the more simple prisms.



Uses of Prisms: 90° Change in Direction

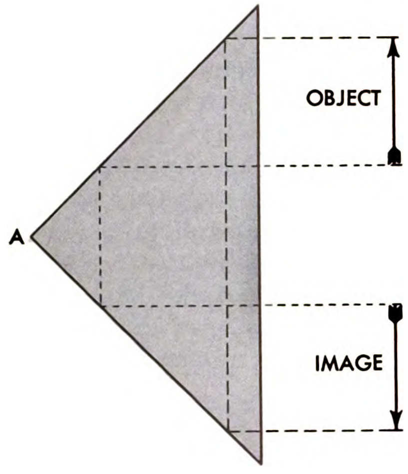
To effect a 90° change in the direction of light, a 45°-45°-90° triangular prism is used as shown in the first diagram. The angle of incidence at the inclined surface, because it is 45°, is greater than the critical angle of the glass of which the prism is made. Hence, total reflection occurs, and all light entering the vertical surface will be transmitted through the horizontal surface.

By placing two 45°-45°-90° triangular prisms in combination as shown in the next diagram, we have a simple periscopic sight. The distance between the horizontal surfaces of the two prisms is of no importance. It is customary to enclose the prisms in a tubular or box-like structure to exclude all extraneous light and to keep dirt and dust off the surfaces of the prisms.



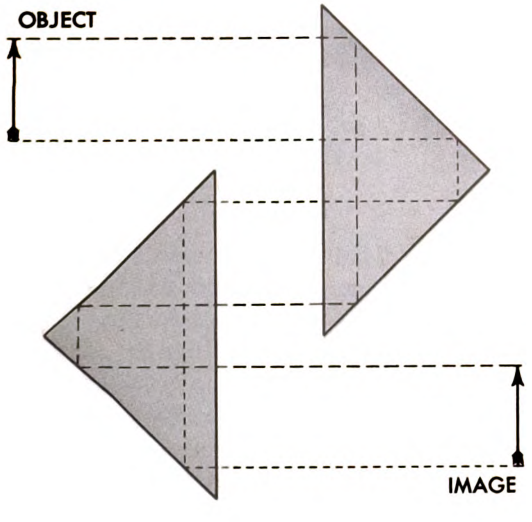
Uses of Prisms: Simple Periscope

To effect a 180° change in the direction of light, a triangular prism is used as shown in the third diagram. Note that the image is inverted. In the design of such a prism, angle A must be 90° to produce the desired effect. This is the only angle that will give the necessary 45° angle of incidence at each of the two reflective surfaces.



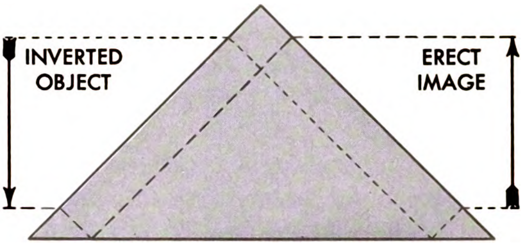
Uses of Prisms: 180° Change in Direction

Prisms of this type may be used to design compactness into optical devices such as telescopic sights and field glasses. In such devices there must be a definite distance between the front and rear lens assemblies. By using a pair of prisms as shown in the next diagram the actual distance between the lenses is shortened to approximately one-third, but the effective distance between the lenses is unchanged.



Uses of Prisms: Double 180° Change in Direction

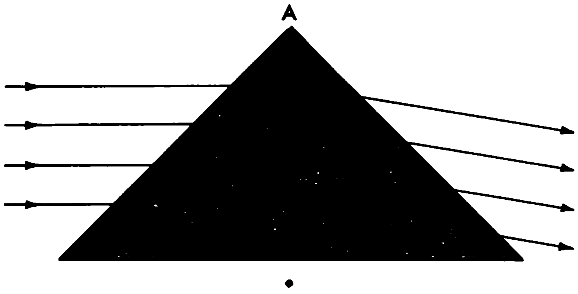
A prism can be used to erect an inverted image by causing the image to be both refracted and reflected within the prism. This is shown in the next diagram. Light composing the inverted image is refracted by the left-hand inclined surface, reflected by the bottom surface, and refracted to its original direction by the right-hand inclined surface. However, the instance of reflection erects or reinverts the image, causing it to appear right-side-up.



Uses of Prisms: Image Erection

REFRACTION BY PRISMS. Except as presented previously, refraction by prisms is not utilized to any great extent in the design of optical sights. Hence, our discussion of this matter at this time is to present the fundamental principles of lenses. Let us look ahead for a moment and see just exactly what a lens is. By definition, it is a piece of transparent material so shaped that the rays of a beam of light passing through it will be caused to converge or diverge. What we are about to demonstrate is that a lens is merely a prism in more complex form.

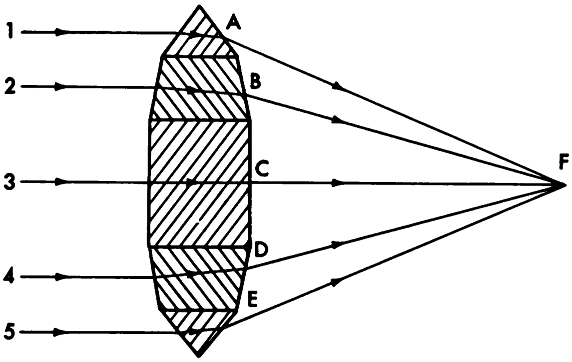
First of all, let us consider a triangular prism through which we pass parallel light rays as shown in the next diagram. Note that after two refractions the rays are still parallel, but that they have been bent downward. This is what we would expect since the angle of incidence is the same at each surface for each ray. The angle at *A* determines the angle through which the rays are bent. This is because a change in the angle of incidence will produce a similar change in the angle of refraction.



Refraction of Parallel Rays by a Prism

If we wish, we may bend the rays of light upward instead of downward. All that is necessary is merely to invert the prism so that angle *A* points downward instead of upward.

Now, let us fashion a crude lens by placing together an assortment of prisms as shown in the illustration. To simplify matters, let us assume that only one light ray reaches each prism and that all the rays are parallel. Since the angle of incidence is zero

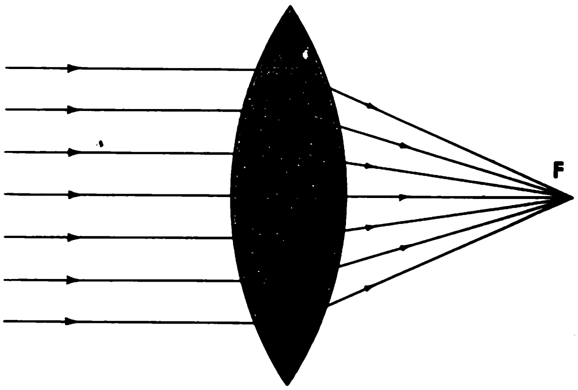


at rectangular prism *C* in the center, ray *3* is not refracted but passes straight through. Small angles of incidence occur at the surfaces of prisms *B* and *D*. Hence, rays *2* and *4* are refracted toward ray *3*. Since the angles of incidence are even greater at the surfaces of prisms *A* and *E*, rays *1* and *5* are refracted even more toward ray *3*. If the angles between the surfaces of prisms *A*, *B*, *D*, and *E* are carefully chosen, all five rays will converge at point *F*.

We must be careful to note that what we have here is not a true lens. Since the surfaces of the individual prisms are flat, all parallel rays reaching each prism remain parallel after refraction. If we were to consider a large number of parallel rays, there could be no true convergence at *F* because, by definition, parallel lines never meet. A true lens must be designed so that, theoretically at least, all parallel lines converge at a point. So let us see what requirements must be fulfilled to produce a true lens.

Lenses

CONVERGING LENSES. From the preceding discussion we can infer that a lens, in order to cause the convergence of a large number of parallel rays, must be shaped so that the angle of incidence varies continuously from zero at the center of the lens to its maximum value near the edge. This can be accomplished by grinding the lens to a spherical contour as shown. If the lens is so shaped,



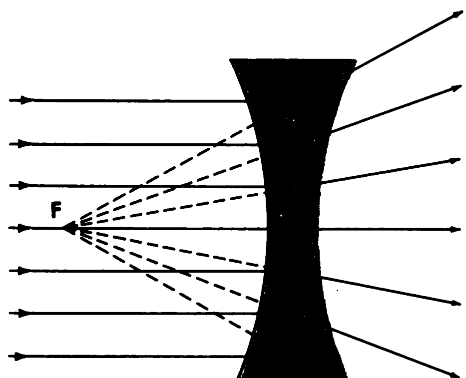
all parallel light rays transmitted through the lens converge at point *F*, the *focus* or *focal point*. The distance between the lens and the focal point is called the *focal distance*. In the case of a thin lens—one whose thickness is small compared to the focal distance—the focal distance may be measured from either surface of the lens.

In the illustration referred to previously, a convention commonly used in optical diagrams should be noted. When dealing with

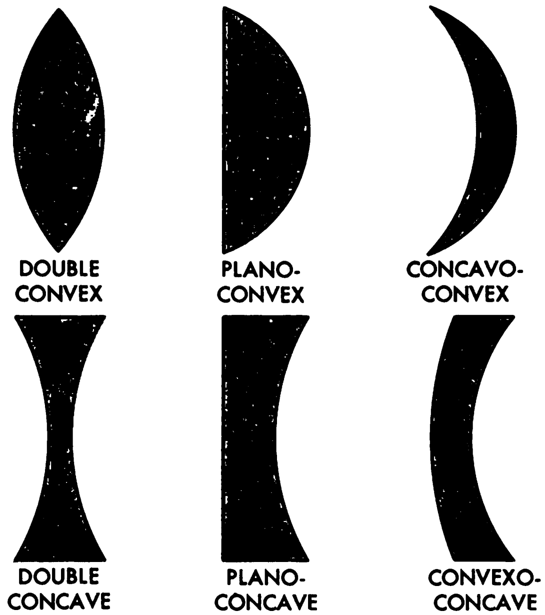
thin lenses, it is often inconvenient to show the refraction that occurs at each surface of the lens. Hence, only one refraction is shown taking place midway between the two surfaces.

DIVERGING LENSES. It has already been pointed out that lenses may cause parallel light rays either to converge or to diverge. So far we have talked almost exclusively about converging lenses. What about diverging lenses?

Going back to our discussion about refraction by simple prisms, you may recall that merely by inverting a prism the parallel light rays will be refracted in the opposite direction. From this we can conclude that, if a convex lens causes parallel rays to converge, a concave lens will cause them to diverge. This is shown in the diagram. Note that the rays appear to diverge from a single point F, which is the focus or focal point of this type of lens.



TYPES OF LENSES. The classification of lenses according to use, converging or diverging, is adequate for general usage. However, a more specific classification results when they are typed according to shape. There are six basic shapes of lenses as illustrated. Reading from left to right, the upper three are: double-convex, plano-convex, and concavo-convex. In the same order the lower three are: double-concave, plano-concave, and convexo-concave. In case you have trouble distinguishing between the third shape in each group, note that maximum thickness occurs in the center of one and at the edge of the other.

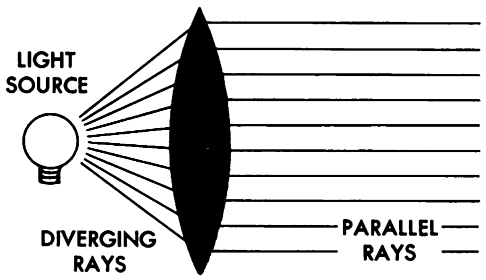


Basic Lens Types

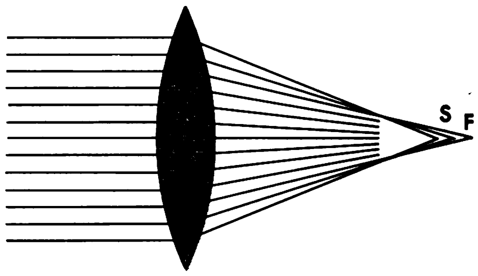
Because they are thickest at the center, all three lenses of the upper group are converging lenses. All three of the lower group are diverging lenses because they are thickest at the edge. The shapes shown in all cases are typical; there are many possible variations of each basic shape.

USES OF LENSES. Lenses are used in optical sighting devices principally as collimators or as condensers.

Collimators are lenses used to change diverging rays of light to parallel rays. Earlier we saw that a double-convex lens would cause parallel rays to converge. This same lens can be used as a simple collimator by operating it backwards. In other words, if a light source such as a lamp is placed at the focal point, the rays transmitted through the lens will be parallel as shown in the diagram.



Lens Used as a Collimator



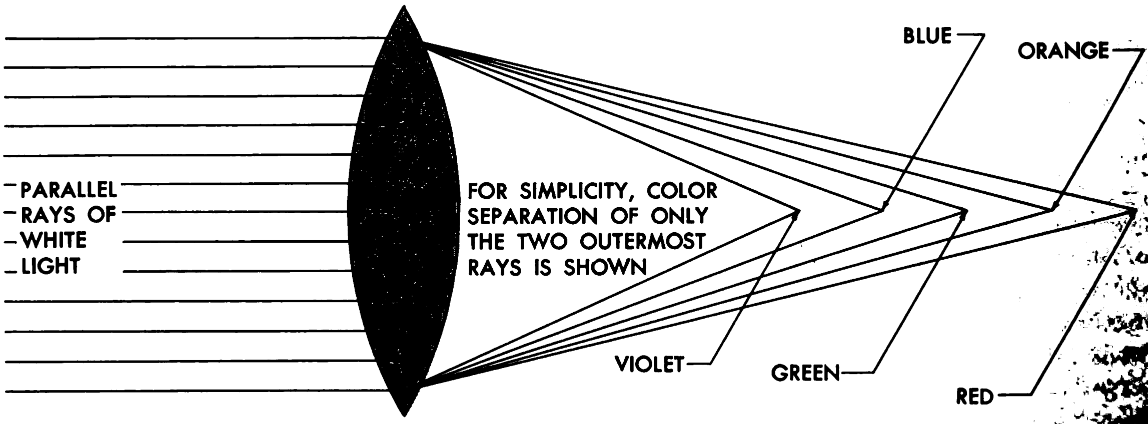
Spherical Abberration

In the interest of simplicity, the diagram shows a single lens used as a collimator. In practice an assembly of two or more lens elements would be used. The additional elements are required to correct for such things as spherical and chromatic aberration.

The curved contour of practically all lenses—regardless of whether they are concave or convex—is spherical in nature. Spherical aberration refers to the fact that parallel light rays passing near the edge of the lens are refracted to a different focal point than those passing closer to the center of the lens. This effect is shown in the diagram. Note that the central rays converge at point *F*, while the outer rays converge at point *S*. This effect is due to the spherical contour of the lens and, therefore, is not caused by imperfections in the lens surface.

The correction of spherical aberration by modifying the contour at the edge of a lens is not as simple as it might seem because of the complications introduced into the lens-grinding processes. Instead, the corrections are made by adding a lens element so shaped that there is little or no refraction of the central rays and a slight divergence of the outer rays.

Chromatic aberration means that each of the various colors that combine to make white light has a different focal length. An exaggerated view of this effect is shown. The basic principle involved is the fact that, because each color has a different speed through a given medium, the angle of refraction varies for each color, even though the angle of incidence is the same for all. Correction for chromatic aberration is obtained by adding a lens element which has



Chromatic Abberation

a color refraction characteristic that is equal and opposite to that of the main lens element.

Condensers are used to obtain brighter images by concentrating rays from a light source into a small area. A condenser usually consists of a pair of plano-convex lenses positioned with their curved surfaces toward each other as shown. The light source *L* is placed at the focal point of lens *1*. The place at which we wish to make use of the concentrated rays may be located anywhere between lens *2* and its focal point *F*. The intensity of the resulting light increases as *F* is approached.

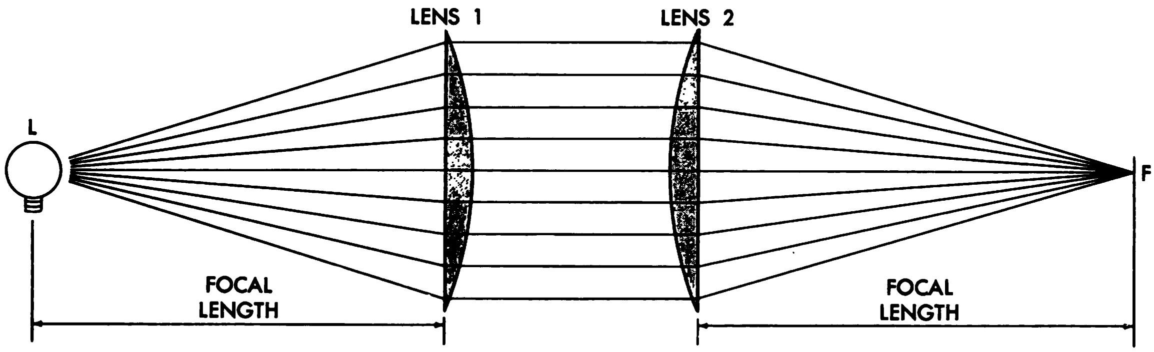
In cases where no correction is desired, the two lenses of a condenser assembly are positioned with an air space between them. In small, compact assemblies where correction of one kind or another is necessary, a double-concave lens of the proper type of glass may be fitted between the two plano-convex lenses.

NOTE

Because the spacing of the elements in a lens assembly is usually quite critical, the disassembly and reassembly of lens assemblies generally should not be attempted at field or organizational maintenance level.

Optical Sights

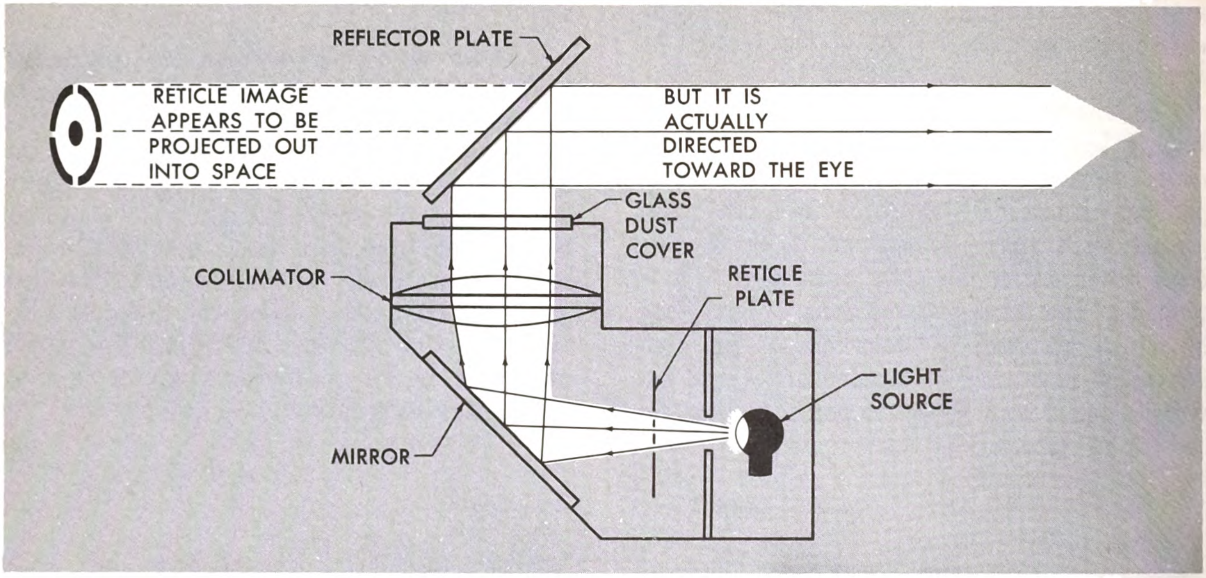
Though various types of optical sights are or have been in use, their basic principles are the same. There is little difference between the basic design of the optical system of the fixed sight and that of the computing sight. The major function of each is to establish the line of sight. In a fixed sight, the line of sight is stationary with respect to the gun line. In a computing sight, the difference between the line of sight and the gun line varies according to the prediction angle derived by the computing system.



Basic Condenser

BASIC FIXED SIGHT. As illustrated, the basic fixed sight has four main parts: light source, reticle plate, lens assembly, and reflector plate. Usually all of these components except the reflector plate are enclosed in a metal housing, which is sealed by a glass dust cover plate for obvious reasons. The reflector plate is attached to the housing by a framework of some type.

Light Source. The light source or sight lamp is usually enclosed in a cavity within the housing. A light-tight access door is provided so that the lamp may be conveniently replaced when necessary. As a safety factor, the lamp is constructed with two filaments, and a selector switch is provided so that the second filament may be used after the first burns out. Except for a small circular area, the glass bulb of some sight lamps is covered



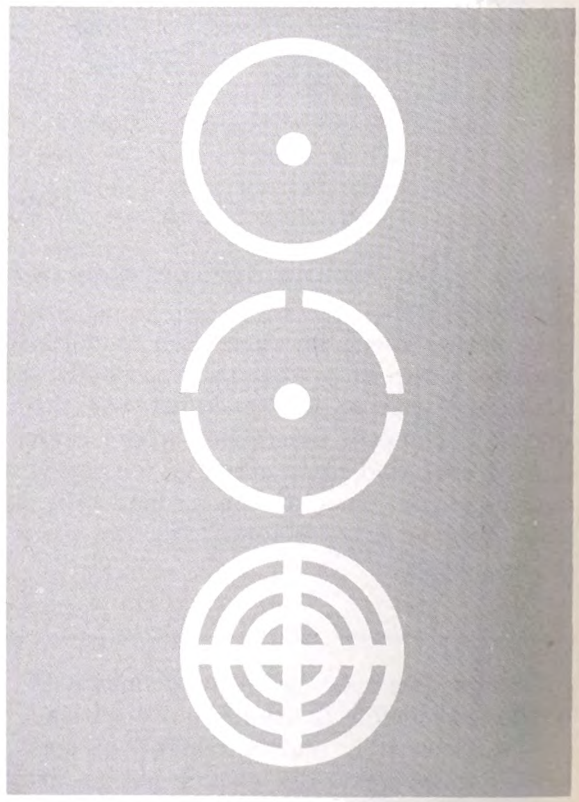
Basic Fixed Sight

with an aluminized reflective coating. This causes practically all the light provided by the lamp to be directed toward the reticle plate.

It is customary to provide the sight lamp with a brightness control so that the amount of light furnished by the source can be varied. This control, which is rheostat connected in series with the lamp filament, may be located in the sight housing or on an external control panel. The use of this control will be discussed later.

Reticle Plate. The reticle plate forms the light pattern or *reticle image* that establishes the line of sight. The reticle plate may be a thin metal diaphragm perforated with the desired pattern. Or it may be a glass plate, one side of which is covered by an opaque coating. In the latter case the desired pattern is etched or engraved into the coating.

Examples of some reticle image patterns of fixed sights are shown. The line of sight is established by the center dot or the intersection of the cross. The circles are used to estimate the desired correction under varying circumstances.



Typical Reticle Image Patterns (Fixed Sight)

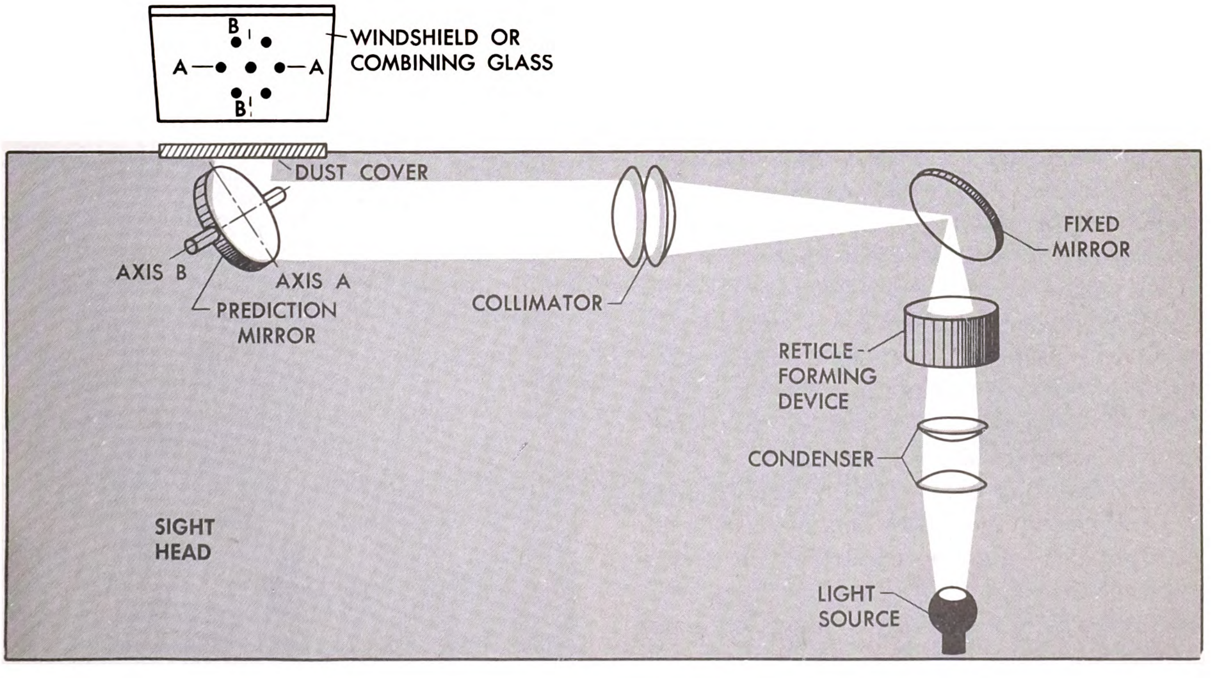
Mirror. The function of the mirror is merely to effect a 90° turn in the direction of the rays coming from the light source through the reticle plate. Some sights use a front-surface mirror for this purpose.

Lens Assembly. The lens assembly is a collimator. The reticle image is formed of diverging light rays which are converted to parallel rays in the process of being transmitted to the reflective plate.

Reflector Plate. The reflector plate is a clear glass plate positioned at an angle of 45° to the parallel light rays transmitted by the collimator lens assembly. The pilot looks through the plate to see the target, and at the same time the plate reflects the reticle image toward his eyes. Since the reticle image is composed of parallel light rays, it appears to be projected out in space and is therefore superimposed on the target image. As a result, the pilot is not required to close one eye and keep his head in a single fixed position as was the case with earlier sights of the ring-and-post or telescopic variety. He is free to move his head in any direction, within limits, without changing the visible relationship between the line of sight and the target.

The brightness control makes possible the use of this type of sight under adverse lighting conditions. When the background is bright, the intensity of the reticle image may be increased; when the background is dark, the reticle image may be dimmed. In any event, the intensity of the reticle image should be kept at the minimum required for satisfactory target observation under the existing conditions. An excessively intense reticle image can conceal the target or blind the pilot with its glare, particularly when light conditions are poor.

BASIC OPTICAL SYSTEM OF TYPICAL COMPUTING SIGHT. The basic optical system of a typical computing sight is shown in the diagram. The sequence of operation begins with



Basic Optical System of a Computing System

the light source at the lower right. The sight lamp is enclosed in a cavity and most of its surface is coated so that practically all the light provided by the lamp is directed toward the condenser lens assembly. The diverging light rays are concentrated into a comparatively small area in passing through the reticle forming device. This makes available a brighter reticle image than would otherwise be possible.

The *reticle forming device* in a computing sight is a rather complicated mechanism and will be described separately later. For the purposes of our present discussion we need only consider that the reticle forming device performs the same basic function as the reticle plate in the simple fixed sight.

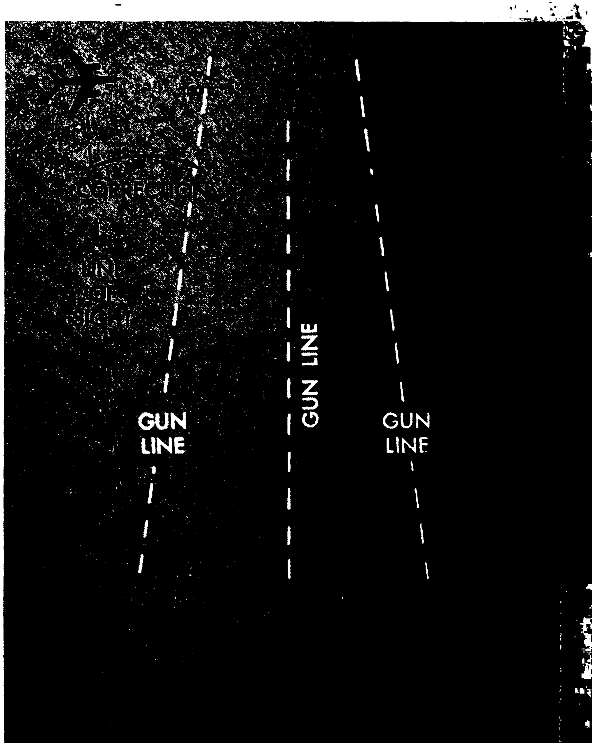
The *fixed mirror* has but one function—to turn the light rays 90° toward the collimating lens assembly. The fixed mirror is located at the focal point of the upper condenser lens. As a result, the converging rays that reach it are reflected as diverging rays toward the collimator lens assembly.

The *collimator lens assembly* performs its normal function of converting diverging rays to parallel rays and transmitting them to the prediction mirror.

The *prediction mirror* has two functions: (1) to turn the light rays approximately 90°, projecting the reticle image onto the windshield or combining glass; and (2) to provide for two-dimensional motion of the reticle image on the windshield or combining glass. In earlier times the aircraft windshield was used in a manner corresponding to that of the reflector plate in the basic fixed sight. This arrangement proved unsatisfactory for various reasons, leading to the use of the combining glass. The combining glass is an optically flat plate mounted between the sight head and the windshield in such a way that the pilot sees through it when looking forward. It is difficult to show, on a flat printed page, the fact that the final turn in the light rays at the windshield or combining glass is actually directed out of the page toward you as you look at the illustration.

Being provided with a gimbal-like mount, the prediction mirror is capable of being rotated about two axes shown as *A* and *B* in the illustration. Rotation about axis *A* causes the reticle image to move from side to side; rotation about axis *B* causes up-and-down motion.

HOW THE PREDICTION ANGLE IS ESTABLISHED. At this point we are in a position to preview the over-all operation of a computing optical sight. The end result of the entire computing system is merely to position the prediction mirror so that the deviation of the reticle image from its neutral or zero-correction position is equal to the prediction angle. The deviation of the reticle image is always in the direction opposite to that of the desired correction. In other words, correction to the right is produced by reticle image motion to the left, and upward correction results from downward motion of the reticle image.

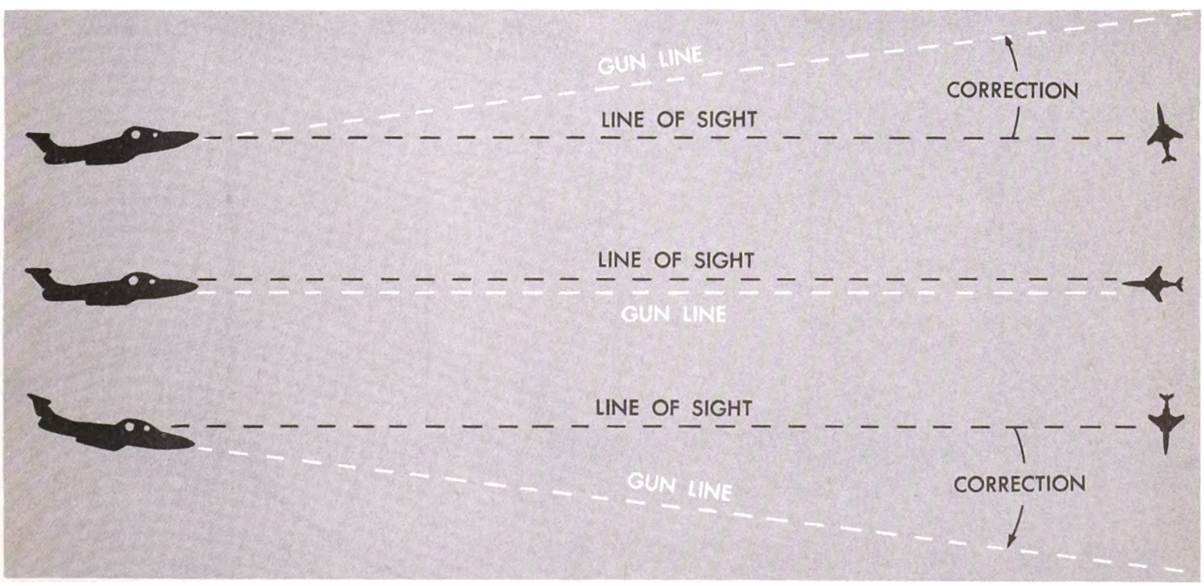


Determination of Correction Angle by Position of Reticle Image (Az)

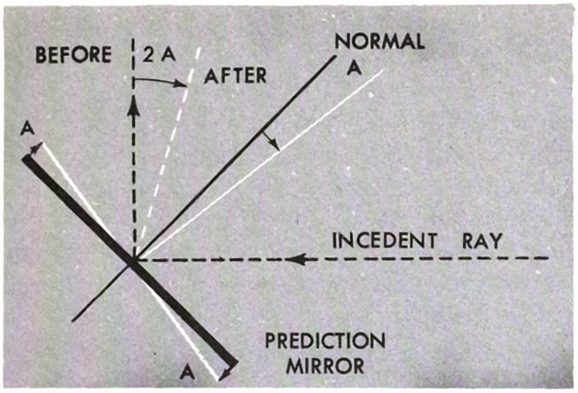
To see how this is accomplished, we must bear two things in mind: (1) the direction of correction is always taken from the line of sight toward the gun line, or from present target position toward future target position; and (2) the pilot flies his aircraft so as to keep the reticle image centered on the target. Now, let us look at the diagram to see how corrections are made in azimuth or deflection. If the reticle image moves to the left and is centered on the target the gun line is automatically moved to the right. If the reticle image is at its neutral position and is centered on the target, the gun line and line of sight coincide—there is no correction. Then, if the reticle image moves to the right and is centered on the target, the gun line is automatically moved to the left.

With respect to corrections in elevation, let us turn our attention to the next illustration. If the reticle image moves downward, the pilot must raise the nose of the aircraft to keep the reticle image centered on the target. This raises the gun line above the line of sight. However, when the reticle image is at its neutral position and is centered on the target, the nose of the aircraft is aimed directly at the target and there is no correction. When both are horizontal, the small vertical difference between line of sight and gun line is disregarded. But if the reticle image moves above the neutral position, the nose of the aircraft must be lowered to center the reticle image on the target. Now the gun line is dropped below the line of sight.

Because the prediction angle is established by a pivoted mirror, the computing system is required to rotate the prediction mirror only through one-half of the desired correction angle. This can be demonstrated by a simple diagram and a bit of mathematical logic. This diagram is quite similar to the one discussed previously. However, in this case we have a mirror which rotates and an incident ray whose direction does not change. Now rotate the mirror through any small angle such as shown at *A*. At the same time the normal will be rotated toward the incident ray through the same angle. Hence, the angle of incidence is reduced by an amount



Determination of Correction Angle by Position of Reticle Image (El.)



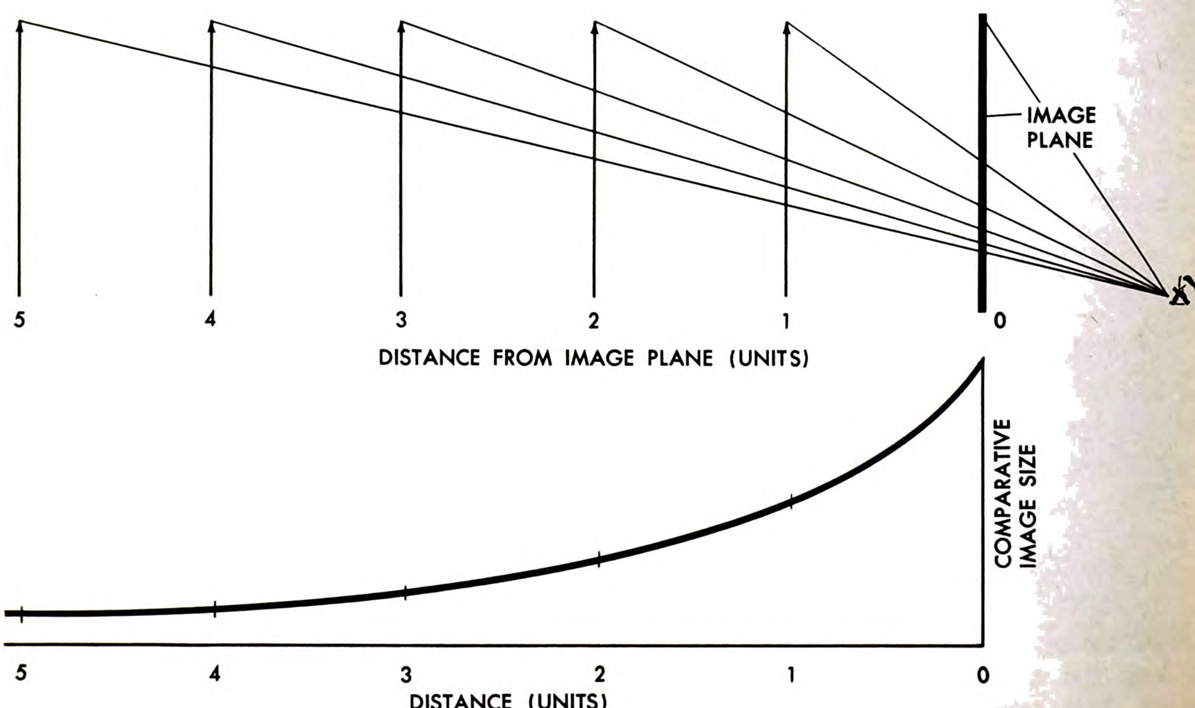
Change in Mirror Position Doubles the Change in Correction

equal to A. Since the angle of reflection is always equal to the angle of incidence, the angle of reflection is also reduced by an amount equal to A. Note that the total angle between the incident ray and the reflected ray has been reduced by an amount equal to two times A. By using the same type of reasoning, you should be able to see that rotating the mirror through angle A in the opposite direction would increase the total angle by two times A. Therefore, the angular change in the direction of the reflected ray is always twice the angular change in the pivoted mirror.

Reticle Forming Device. In a computing sight the reticle image consists of a circle of illuminated spots with an additional spot in the center. To provide for sight operation under emergency conditions, a rather complex reticle forming device is used to make possible the variation in the diameter of the circle. This arrangement permits the reticle image to be used for manual ranging or to be set up as a standard fixed-diameter circle for use as a ring-and-post sight.

Under normal conditions, a servo loop causes the circle diameter to vary according to the output of the radar ranging system, giving an indication of whether the radar system is functioning properly.

The method used for manual ranging is called stadiametric ranging. The basic principles involved in this method have been



Effect of Distance on Image Size

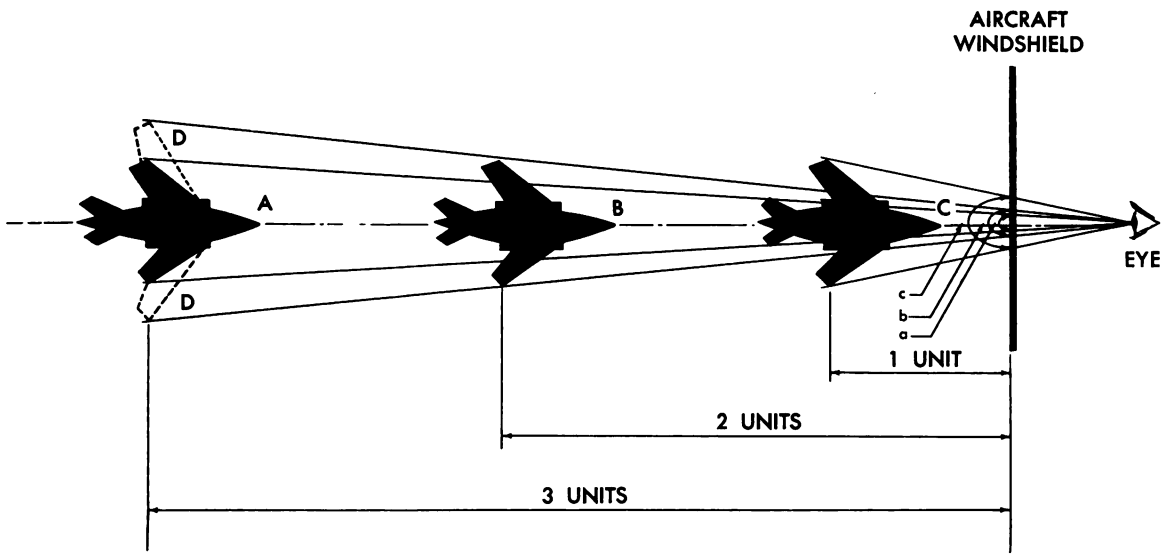
utilized in surveying for a number of years, so we are dealing with something that isn't particularly new.

Stadiametric Ranging. When you look at an object, an image of the object is formed in your optic system. The size of the image you see depends on two factors, the size of the object and the distance between you and it. The size of the image is directly proportional to the size of the object. However, the relationship between image size and distance is not quite so simple. This is shown graphically in the accompanying diagrams. In the upper portion we see an object represented by an arrow at various distances from the eye. For any distance the image size is equal to the intercept on the image plane.

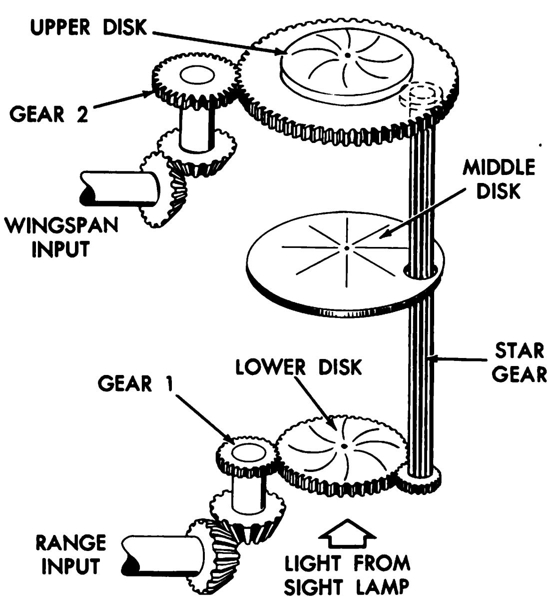
In the lower portion we plot image size against distance from the image plane. Note that the resultant curve is not a straight line but a spiral. Hence, the change in image size is not directly proportional to the change in distance. However, it must be admitted that the effect is shown in exaggerated form. In actual practice, the distance between the eye and the image plane is quite small compared to the distance between the image plane and the object. This causes the curve to approach a straight line more closely than is shown.

Now let us see how this principle applies to target ranging. In the next diagram, let us assume that a target aircraft is flying directly toward us. We first see the target at position *A*. The range is three units, and the image dimension is represented by dimension *a* on the windshield. A moment or so later, the target is at position *B*. The range has decreased to two units, and the image size has increased to dimension *b*. Finally, the target moves to position *C*. This decreases the range to one unit, and increases the image size to dimension *c*.

Now let us go back to position *A* and assume that we have a larger target at a range of three units. The wingspan of the new target is indicated by the dashed wing outline *D-D*. This new target could give us the same image size as the original target at position *B* where the range is two units. Therefore, we have to take target size into consideration before we can use image size to obtain range information. Manual ranging is accomplished when the size of the reticle image, compensated to target wingspan, is made equal to the wingspan of the target image.



Mechanical Details. The reticle forming device usually consists of three patterned discs and a planetary gearing system as shown in the diagram. The three discs are arranged so that light passing through them in succession produces a circle of illuminated spots as well as a center dot. The diameter of the circle can be changed by varying the inputs to the mechanism.



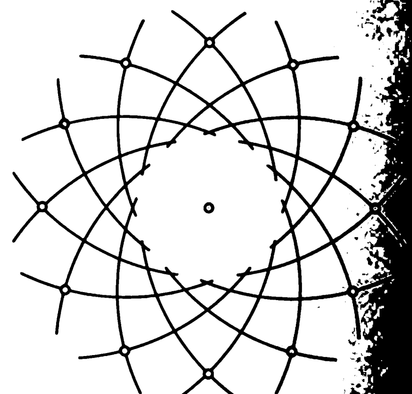
Almost any number of spots can be used to form the circular patterns, but the preferred number is ten. This number of spiral slots is cut into the upper and lower discs. Note that they curve in opposite directions. The same number of straight slots is cut into the middle disc. A small hole in the center of each of the three discs produces the center dot. Wherever the corresponding slots of the three discs have a common intersection, a thin beam of light passes through. There being ten such intersections at all times, ten illuminated spots arranged in a circle complete the reticle image.

The diameter of the reticle image circle is determined by the relative positions of the three discs. Changing their positions causes the intersections to move toward or away from the center dot. Here is how it is done:

The three discs are connected in a planetary gearing system. The output of the wingspan lever is coupled to the rim of the upper disc. The range input is coupled to the rim of the lower disc. The middle disc is geared to the other two discs in such a way that it follows in the same direction at half speed when either of the other two discs turns.

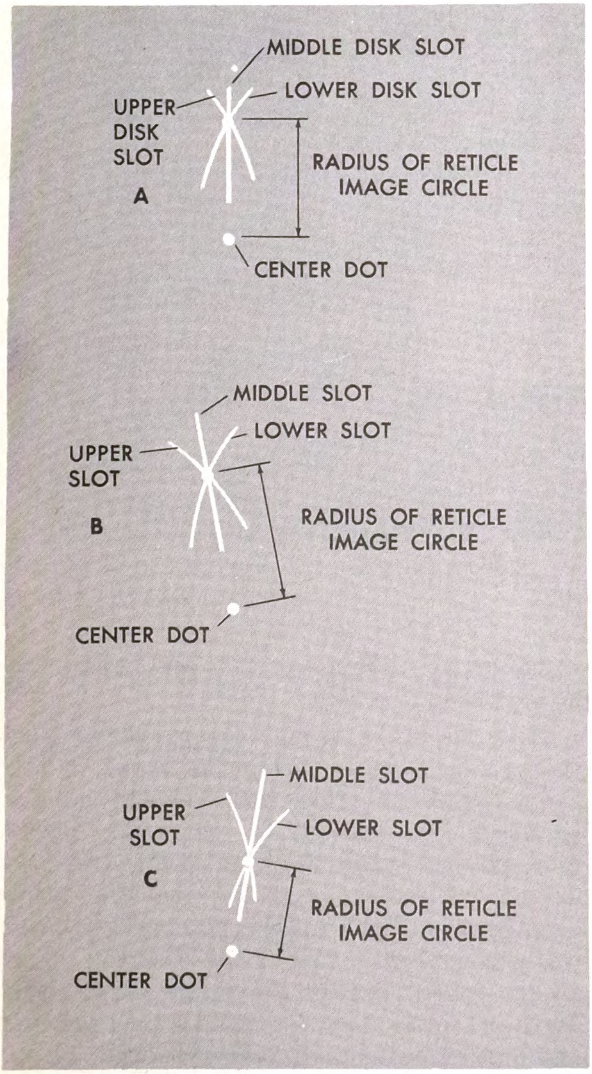
Let us first take an example where the changing range input causes gear 1 to rotate clockwise. This causes the lower disc to rotate counterclockwise. The motion of the lower disc cannot rotate the star gear without causing it to revolve along the internal gearing of the upper disc. (The star gear is actually a single gear attached to the middle disc and meshed with both the internal gearing of the upper disc and the external gearing of the lower disc.) The rotation of the star gear about its own axis is clockwise, but its revolution about the internal gearing of the upper disc is counterclockwise. This revolution of the star gear causes the middle disc to rotate in the same direction, counterclockwise. Checking back, we note that the middle disc rotated in the same direction as the lower disc. By using a similar follow-through, we would find that the middle disc will follow in the same direction whenever the upper disc rotates.

In the next diagram, the 30 slots of the three discs are shown superimposed to illustrate how the circle of ten spots is formed about the center dot.



To see how the diameter of the reticle image circle is changed, let us consider the accompanying series of diagrams. In these we show only one of the ten sets of superimposed slots and the center dot. In the top diagram, the radius of the circle is set up for some definite range and wingspan dimension. The middle diagram shows the effect of a decreasing range input. The lower and middle discs are rotated counterclockwise, while the upper disc remains stationary. The result is an increase in the circle radius.

The bottom diagram shows the effect of an increasing range input. Now the lower and middle discs rotate clockwise from their original position in the top diagram, while the upper disc again remains stationary. The result is a decrease in the circle radius.



Changing the wingspan input will have a similar effect on the size of the reticle image circle. However, in this case the upper and middle discs rotate, and the lower disc is stationary. From a tactical point of view, there is a distinct difference. The wingspan input is generally set for a specific target dimension and left that way as long as the type of target aircraft remains the same. Hence, the wingspan input establishes the basic diameter of the circle, and the range input varies the diameter between its maximum and minimum limits.

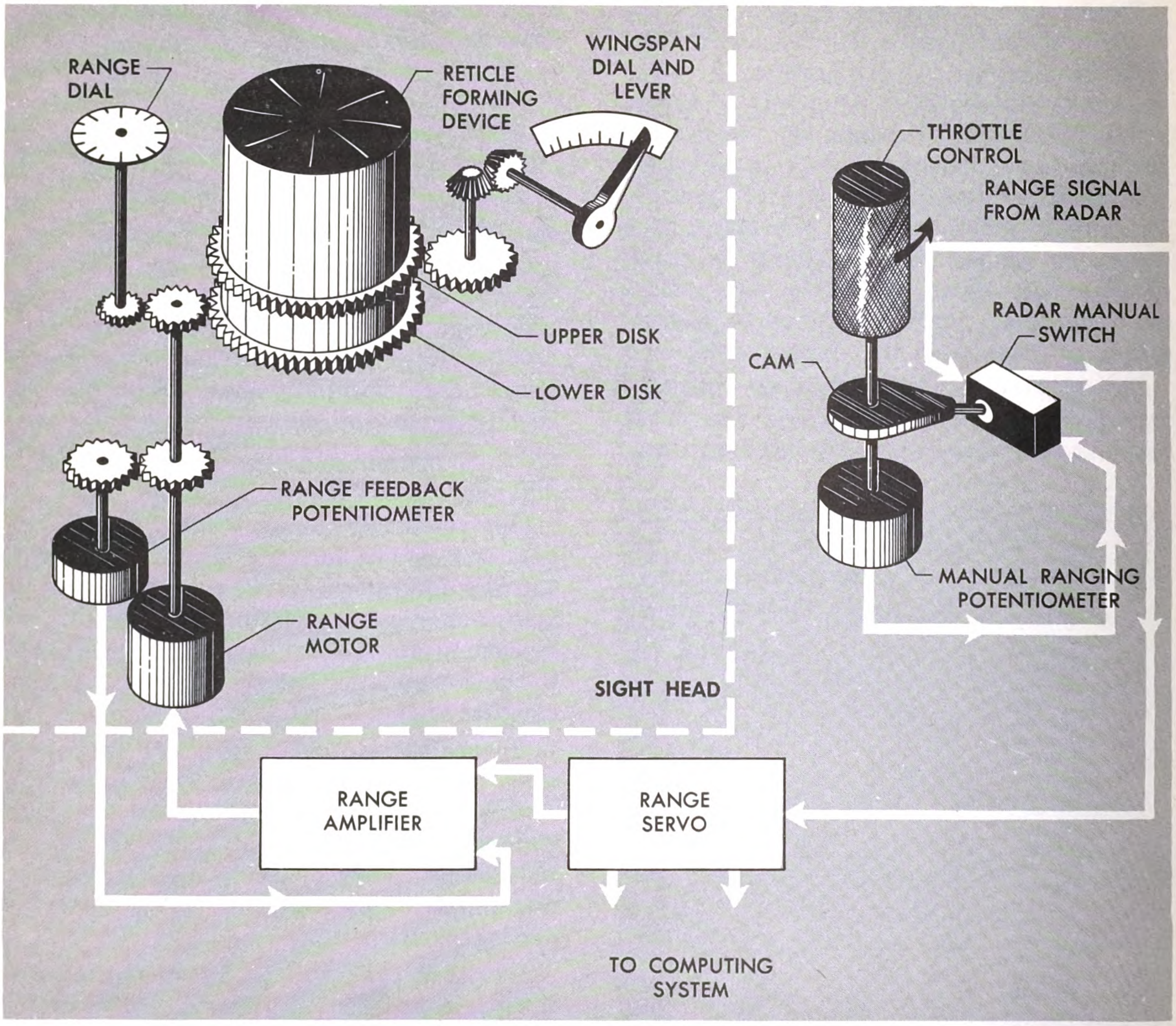
Operation. For our discussion of the operation of a reticle forming device, let us consult the diagram. Here we combine a simplified mechanical schematic with an electrical block diagram. The system shown is designed to accommodate either automatic radar ranging or manual ranging.

Regardless of which type of ranging is used, the wingspan lever is used to set the anticipated target dimension on the dial. This operation establishes the basic size of the reticle image circle by positioning the upper disc of the reticle forming device as previously described.

In *radar ranging,* the twist grip of the throttle control is left in its detent position. In this position the cam depresses the plunger of the radar-manual switch, causing the output of the radar ranging system to be connected directly to the range servo.

The range servo is a vastly more complex device than a simple block diagram indicates. It has several functions, which basically consist of adapting the computing sight system to the particular radar set used for ranging and to the particular weapons with which the aircraft is equipped. The main point is that the electrical output of the radar set is not directly usable in the computing system.

The range servo converts the radar signal into three distinct electrical outputs. One of them whose value is purely a matter of



Manual Ranging

range is transmitted to the range amplifier. The other two are combined with other inputs to the sighting system and transmitted to the computing system.

The electrical output of the range amplifier drives the range motor which, in turn, drives the lower disc of the reticle forming device, the range dial, and the range feedback potentiometer. The output of this potentiometer is applied to the input circuit of the range amplifier. Hence, the actual input to the amplifier is the algebraic sum of the feedback signal and the signal from the range servo. When the two are equal, the effective input is zero.

Here we see a remote-control servo loop in its most simple form. The input to the range amplifier from the range servo determines the position to which the lower disc of the reticle forming device will be driven by the range motor, and the feedback input to the amplifier determines when the positioning of the lower disc has been accomplished. Hence, the range motor will drive the lower disc to a definite position for any given range signal from the radar. Thus,

the diameter of the reticle image circle and the indication on the range dial are determined by the output of the radar system.

In *manual ranging,* the twist grip of the throttle is rotated away from its detent position. This causes two things to happen: (1) the turning of the cam releases the plunger of the radar-manual switch, disconnecting the radar signal and connecting the output of the manual ranging potentiometer to the range servo; and (2) the electrical output of this potentiometer varies with the degree of rotation of the twist grip. Except for the change in the signal source, operation is basically the same as for radar ranging. Now the size of the reticle image circle changes according to the output of the manual range potentiometer. When the grip is rotated so that the diameter of the reticle image circle is equal to the target wingspan —and kept that way—the range servo supplies the proper range signals to the computing system.

As mentioned previously, manual ranging operation is provided to permit functioning of the computing system under emergency conditions. No completely satisfactory system for manual ranging as applied to fighter/interceptor aircraft has ever been devised. This is the principle reason why automatic radar ranging has been developed and is used so extensively. However, it must be admitted that, in the event of a radar malfunction, any manual ranging system is better than no ranging at all.

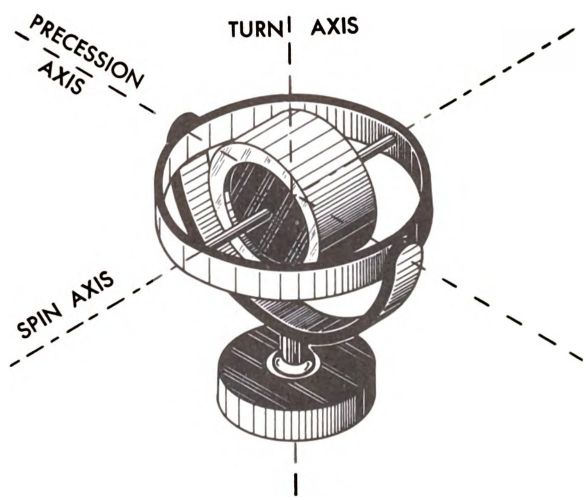
USE OF GYROSCOPES IN COMPUTING SIGHTS

Gyroscopic Fundamentals

A gyroscope, which we will call a *gyro* for short, is a mechanical device whose physical appearance is easily described. Its principles of operation, however, are an entirely different matter. To explain fully *how* a gyro works involves our getting into physics and mathematics to an extent that is beyond the scope of this manual. Therefore, we will confine our discussions to what a gyro looks like, what it does, and why it is necessary in a computing sight.

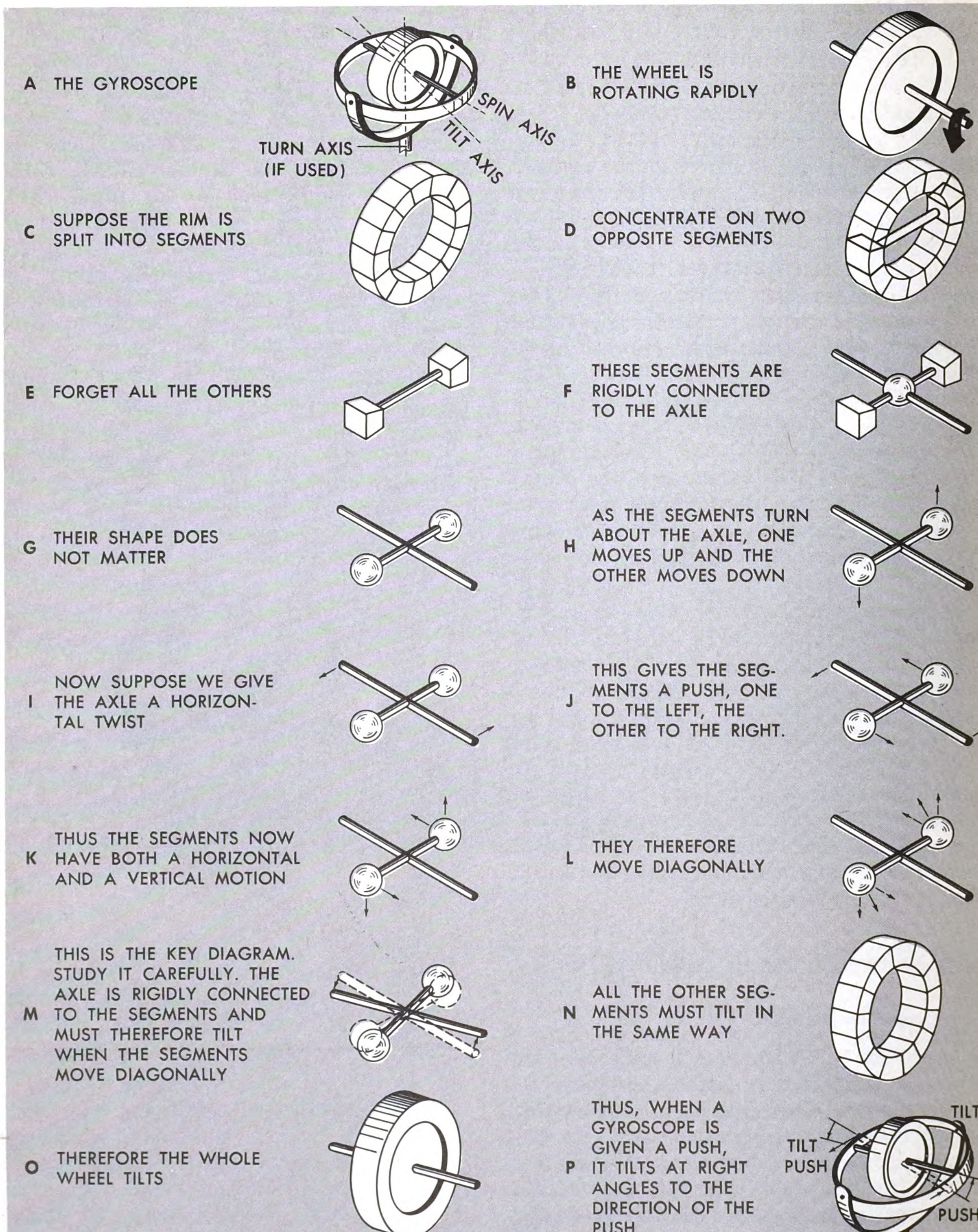
DESCRIPTION. There are various types of gyros, but the basic component of all of them is an accurately balanced flywheel spinning at a constant high speed. This flywheel rotates about its central axis, called the *spin axis,* which passes through the center of gravity of the wheel. Various methods are used to drive the wheel. The only important consideration is that, whatever method is used, it should be capable of maintaining the desired spin rate within fairly close limits.

The basic type of gyro is mounted on gimbals (ring mounting) so that the wheel is free to turn or tilt in any direction about its center of gravity. This type of mount is shown in the illustration.



Two special characteristics make gyros valuable for use in conjunction with computing systems. These characteristics are: (1) gyroscopic inertia or rigidity in space, and (2) precession.

RIGIDITY IN SPACE. Rigidity is the tendency of a gyro to hold its spin axis at a fixed direction in space and to resist any forces which tend to displace or change in direction of the spin axis. A gyro wheel, spinning at high speed and supported by a gimbal mount as illustrated previously, will maintain its



Gyroscopic Precession

spin axis in the same direction unless some outside force is applied to it.

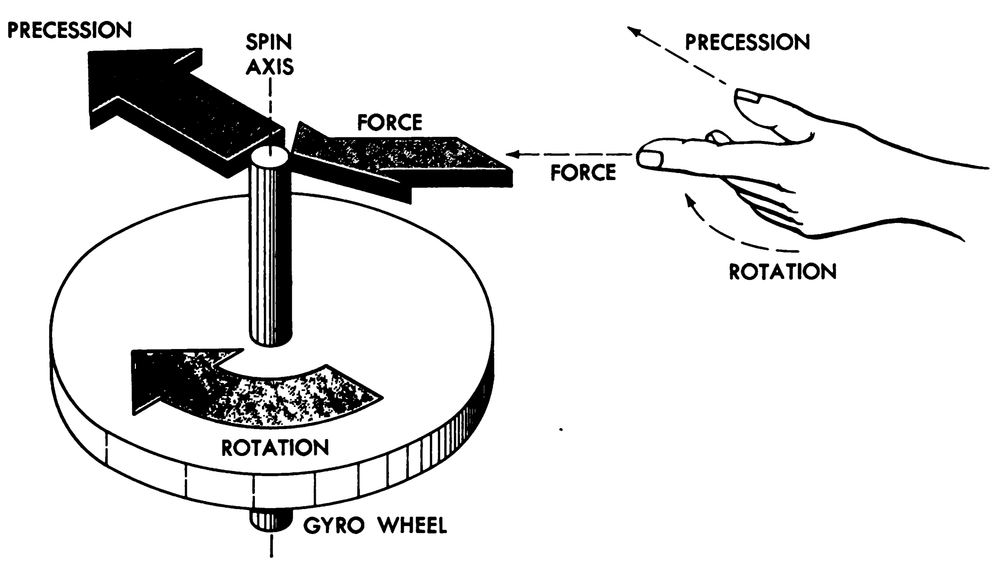
Three factors determine the amount of rigidity in a particular gyro: the weight of the wheel, the distribution of this weight, and the speed at which the wheel rotates. Let us consider these factors one at a time and see how each affects the rigidity of a gyro.

The ability of a gyro to maintain its position in space can be increased by using a heavier wheel. A gyro with a heavy wheel will have more rigidity than one with a light wheel spinning at the same rate. Further, for any given wheel weight, the rigidity can be increased by massing the weight as far from the spin axis as possible. This is why the wheels of small gyros are shaped like a cup. The distribution of the bulk of the weight to the outer rim produces the same amount of rigidity as a larger wheel of the same weight rotating at the same speed. Finally, the rigidity of a gyro is increased with an increase in the speed at which the wheel rotates. A slowly spinning wheel gives little or no rigidity.

Computing systems involving gyros are designed with a fixed amount of rigidity in mind. For this reason, whatever method of propulsion is used to rotate the gyro wheel must include provisions for maintaining the spin rate within fairly close limits. When AC power is used, the gyro motor is usually of the synchronous type, and speed regulation is obtained by keeping the frequency of the AC within close tolerance. When DC power is used, the gyro motor is usually provided with a speed governor of some type.

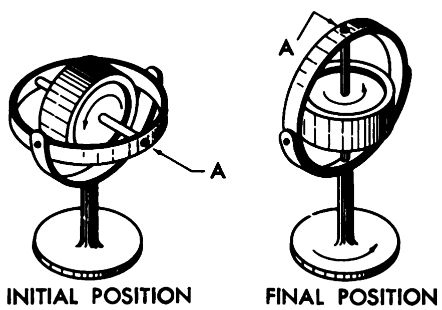
PRECESSION. The rigidity of a gyro resists any applied force which would turn the spin axis in any direction. When enough turning force is applied to overcome the rigidity, the spin axis will turn—but not in the direction of the applied force as you might expect. Instead, the spin axis turns at right angles to the applied force and in the direction of gyro wheel rotation. This peculiar action is the second of the gyro's two special characteristics and is called *precession.* The picture story shown is the simplest explanation of why a gyro precesses that it is possible to give you.

A simple procedure will enable you to determine the direction of precession for any gyro. Extend the thumb and index finger of *either* hand to form an approximate right angle. Let the remaining fingers curve naturally. (The thumb and the curved



Direction of Precession

fingers should point in the same general direction.) For any gyro whose wheel is spinning in the direction indicated by the curved fingers, a force applied to a point on the spin axis in the direction of the index finger will cause the same point to precess in the direction of the thumb as shown in the illustration. In making use of this procedure, use whichever hand whose fingers conveniently indicate wheel rotation. It will be noted that one hand, when inverted, gives exactly the same results as the other hand in its normal attitude.



Extent of Precession

The precessional movement of the gyro continues as long as the force is applied to the spin axis, but it stops when the spin axis takes a position where the direction of the applied force is the same as the direction of wheel rotation. This is shown in the illustrations. Assume that the gyroscope at the left is rotated continuously in a counterclockwise direction and that a force is applied at point *A* toward your right and away from you. Using the hand procedure, you can quickly determine that point *A* will precess upward. Precession continues until point *A* is at the upper extremity of the wheel shaft as shown at the right. Now the applied force is in the same direction as the rotation of the wheel. Further rotation of the entire gyro cannot produce precession because the applied force no longer changes the direction of the spin axis.

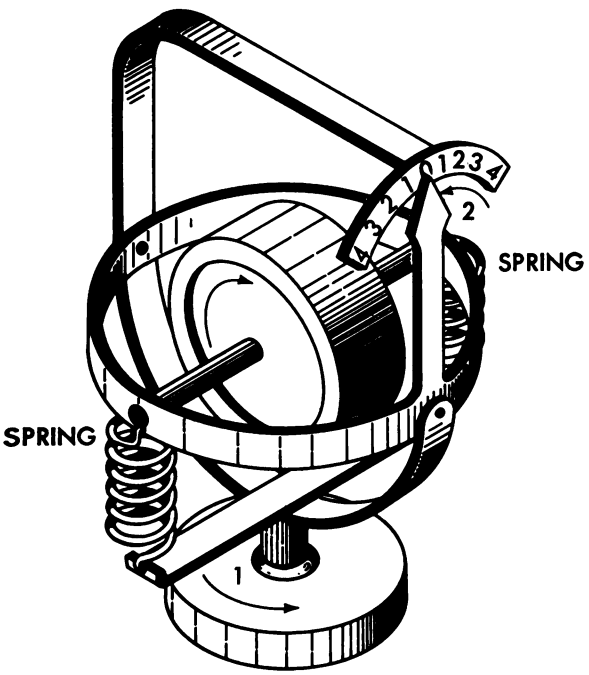
If the gyro is rotated in the direction opposite to that shown, precession will occur until the wheel shaft is vertical with point *A* at its lower extremity. In this attitude the direction of wheel rotation is the same as that of the opposite turning force. There is no further precession because the turning force again no longer disturbs the position of the spin axis.

RATE GYRO. A rate gyro is one that produces a precessional force that is proportional to the rate at which a turning force displaces the gyro spin axis. To understand the basic principles of a rate gyro, let us return to the previous illustration for a moment or two. Suppose that we add mechanical stops in such a way that the gyro is permitted to precess only a small portion of the way up or down toward the vertical position of the spin axis. Then, if we impart continuous rotation to the gyro, we will first observe precession to the mechanical stop. Following this, the gyro will bear against the stop with a force that is proportional to the rigidity of the gyro and the rate at which the turning force is applied to the gyro. (In other words, the gyro will try to overcome any restraint that prevents it from precessing all the way.) If the speed of the gyro wheel is constant, the amount of rigidity is constant. When rigidity is eliminated as a variable, the force exerted against the stop is proportional to the rate of the applied turning force.

While a gyro equipped with mechanical stops is useful in showing fundamental principles involved in a rate gyro, it is not usable in a computing system. Here we must have a rate gyro that has the capability of indicating a continuously variable precessional force that is a measure of any applied turning force. This can be done by restraining the gyro in a different manner. For instance, suppose that instead of a rigid mechanical stop we were to use something of an elastic nature such as a rubber band to oppose or restrain gyro precession. It takes force to stretch a rubber band, and the distance that it stretches is a good indication of how much force you are applying. If a gyro is made to work against an elastic force

that opposes precessional movement, the actual amount of precessional movement will be proportional to the turning force applied to the gyro.

A simple but more practical application of elastic restraint to a gyro is shown in the illustration. Here we see two coil springs used to furnish the elastic restraint and to establish the normal attitude of a basic rate gyro. In this example the disturbing force is applied as rotation of the base of the gyro. If no disturbing force is applied, the two springs exert equal and opposite forces on the gyro, and the pointer indicates zero on the scale. If the base of the gyro is turned in the direction of arrow 1, the gyro will precess in the direction that will move the pointer toward the left-hand end of the scale as indicated by arrow 2. (You can check this by applying the hand procedure.) The precession of the gyro stretches the right-hand spring. The pointer will move to the left until the precessional force is balanced by the increase in the tension of that spring.



Basic Rate Gyro

The reading on the scale depends on the speed at which the base is rotated. If the base is rotated at a faster speed, the precessional force is greater. The right-hand spring will have to stretch more to produce enough tension to balance the increased precessional force. As a result, the pointer will indicate a higher reading.

If the gyro is rotated in the opposite direction, precession will also be in the opposite direction. In this case, the left-hand spring will stretch to provide the balancing tension, and the pointer will indicate a reading on the right-hand side of the scale. The readings will still be proportional to the speed of rotation of the gyro.

A rate gyro equipped with scale and pointer such as we have just discussed would find little use in a computing system. Therefore, we must be prepared to consider rate gyros in which the pointer is replaced by a lever. This lever is coupled to a shaft through a crank in such a way that gyro precessional movement is translated into corresponding shaft rotation. Further, we must anticipate some other method of providing elastic restraint because—after all —an arrangement consisting of rubber bands or coil springs is hardly practical.

GYROSCOPIC ERROR. The spin axis of a gyro is not always in the direction in which it theoretically should point. Random inaccuracies in the gyro produce this change in the direction of the spin axis, called *gyroscopic error* or *drift*. There are three general sources of drift.

Unbalance. A gyro can become dynamically unbalanced when operated at a speed or temperature other than that for which it was designed. Some unbalance exists in any gyro since manufacturing tolerances preclude perfect symmetry. Unbalance can be minimized by maintaining a constant spin speed and by operating the gyro in a temperature-controlled area.

Bearing Friction. Friction in the bearings of the gimbals results in lost energy and incorrect gimbal positioning. Friction in the

spin axis bearings causes drift only if the friction is not symmetrical. An even amount of friction all around the bearings results only in a reduction in the spin speed.

Gimbal Inertia. Whenever a disturbing force displaces the gyro spin axis, part of the force that tends to maintain the position of the axis is used up or lost in overcoming the inertia of that part of the gimbal mount that moves along with the wheel. The greater the mass of the gimbal, the greater the resulting drift.

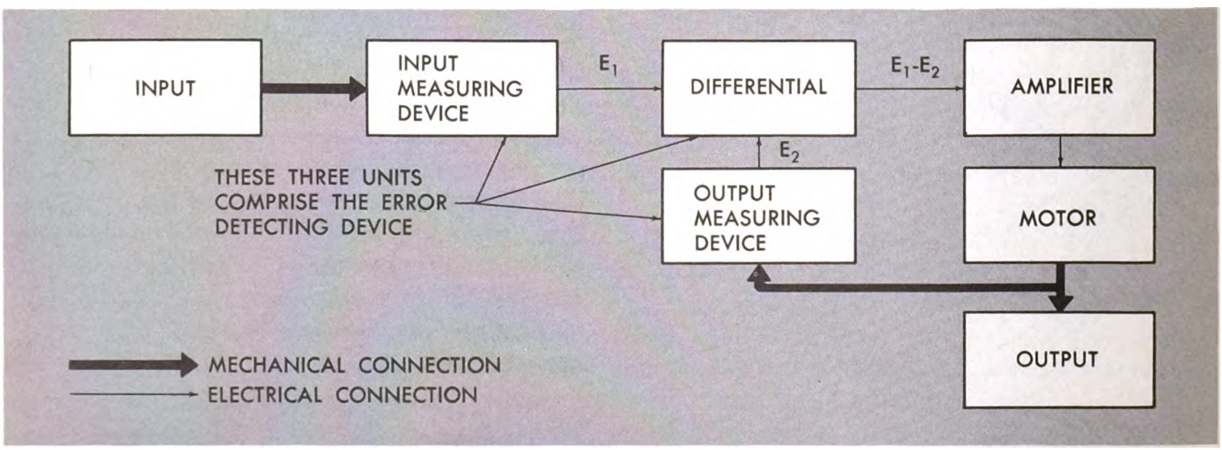
The complete elimination of drift through improved gyro design alone seems to be an impossibility although much progress toward this end has been made in recent years. To date the most satisfactory solution to the drift problem has been to add an erecting or caging device to the gyro as an auxiliary mechanism. Erecting or caging devices eliminate drift by mechanically driving the gyro spin axis to its proper position.

Application of Gyros in Servo Loops

BASIC SERVO LOOP. The primary function of a servo loop is to transmit mechanical position from one place to another by electrical means. In one form or another, servo loops are used in practically all remote control systems. Because of the many possible variations, we will discuss servo loops in the most general terms possible.

Stripped down to its barest essentials, a servo loop consists of an input shaft, an output shaft driven by a motor, an error-detecting device, and an amplifier. When the input shaft and the output shaft are not exactly in the same position, an *error* is said to exist and an *error signal* appears in the error-detecting device. This error signal is amplified so that there is sufficient power to drive the output motor to aline the two shafts. Amplification makes possible the controlling of a large amount of output power with a small amount of input power. Often a servo loop will be used solely for the power boost it affords. This is especially true when sensitivity and accuracy demand that the input shaft carry as light a mechanical load as possible.

In the diagram shown, the input and output of the basic servo loop are represented as mechanical shaft rotations, and the remaining elements are electrical. The input measuring device converts the mechanical input into an electrical signal, E_1, indicative of the input shaft position. The output measuring device converts the mechanical output of the motor into an electrical signal, E_2, indicative of the output shaft position. The differential measures the difference between the two signals, E_1 and E_2. Any difference is amplified by the amplifier and used to drive the motor.



If the output shaft position corresponds to the input shaft position, E_1 and E_2 are equal. Since their difference is zero, the output of the differential is zero. The input to the amplifier is therefore zero, and so is its output. No driving power is applied to the motor, and the output shaft does not move. Thus, when the input and output shafts are in correspondence, the error signal is zero and the servo loop may be considered at rest.

If the input shaft is rotated from this position of correspondence, the voltage put out by the input measuring device changes, increasing in one direction of rotation or decreasing in the other. Since this voltage is now different from the voltage of the output measuring device, the difference is detected in the differential and then is amplified in the amplifier. The resulting output is applied to the motor which drives both the output shaft and the output measuring device.

Since the input shaft can be rotated in either direction, the motor must be reversible, and furthermore the amplifier must be capable of furnishing an output that causes the motor to operate in the direction corresponding to the rotation of the input shaft. The servo loop is designed so that the error voltage is applied in a direction to drive the output shaft into correspondence with the input shaft. As the output shaft moves toward the position of correspondence, the voltage of the output measuring device changes toward equality with the voltage of the input measuring device. The error signal therefore becomes less. When the output shaft reaches correspondence with the input shaft, the voltages of both measuring devices are equal, and the error signal is zero.

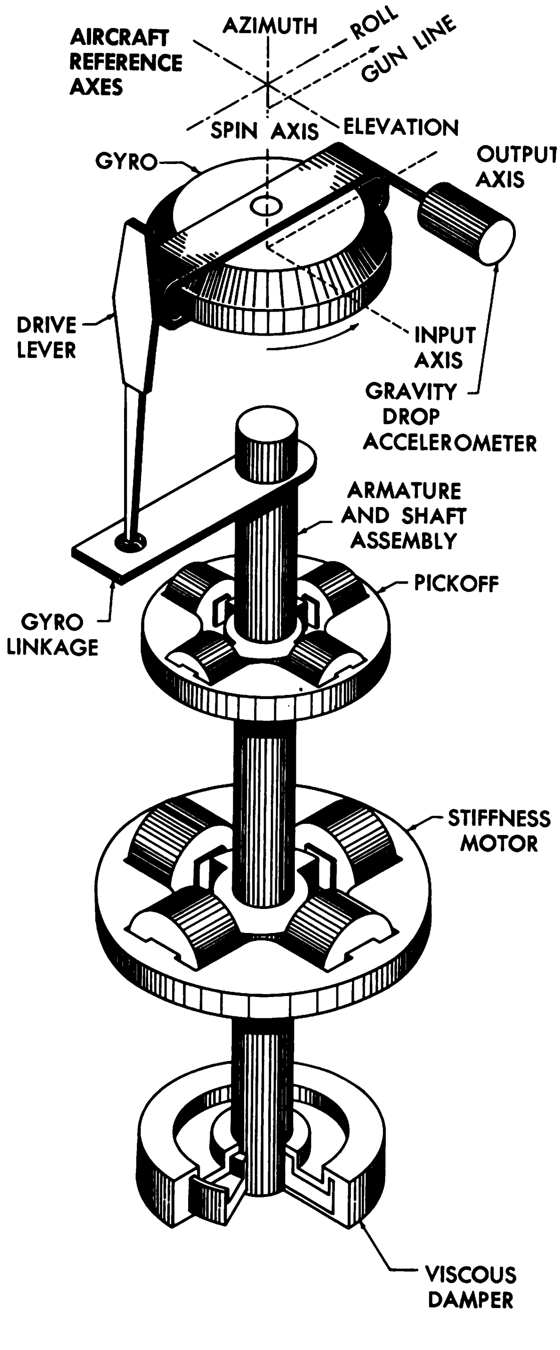
As the motor drives the output shaft toward correspondence with the input shaft, the momentum built up in the motor may cause the output shaft to overshoot the position of correspondence. This causes an opposite error signal, thus reversing the motor and possibly driving the output shaft through the position of correspondence again. A succession of reversals centered about the position of correspondence, called *hunting*, may occur unless provisions are made to eliminate or suppress it. An additional requirement of the servo loops used in fire control systems is that they be designed to be sensitive to as small error signals as possible. In gunnery, correction angles are measured in terms of mils, one of which is equal to approximately one-eighteenth of one degree.

THE COMPUTER ASSEMBLY. The computer assembly of an optical sight consists of gyro, pickoff, stiffness motor, and viscous damper combined into a single compact assembly. It will help us to understand the function of the computer assembly better if we bear in mind that the gyro and the pickoff are the input and input measuring device of the basic servo loop which we have just discussed.

A simplified mechanical schematic diagram of an elevation computer assembly is shown next. Note that the gyro is coupled to the shaft through a lever-and-crank linkage and that the shaft is common to the armatures or rotors of the pickoff, stiffness motor, and viscous damper.

Gyro. Note that the input axis of the elevation gyro is parallel to the elevation axis of the aircraft. Whenever the nose of an aircraft is raised or lowered, rotation about the elevation axis results. Any such rotation displaces the spin axis of the elevation gyro and causes it to precess about its output axis which is parallel to the gun line. The precessional force is proportional to the rate of rotation about the elevation axis, and the direction of precession depends on the direction of rotation about the axis. In other words, the gyro precesses in one direction when the nose of the aircraft is raised, in the other direction when the nose is lowered. When the aircraft travels in level flight, there is no precessional force, and the gyro assumes the position of zero correction.

The mechanical linkage translates the precession of the gyro into rotation of the common shaft. For instance, if the gyro spin axis

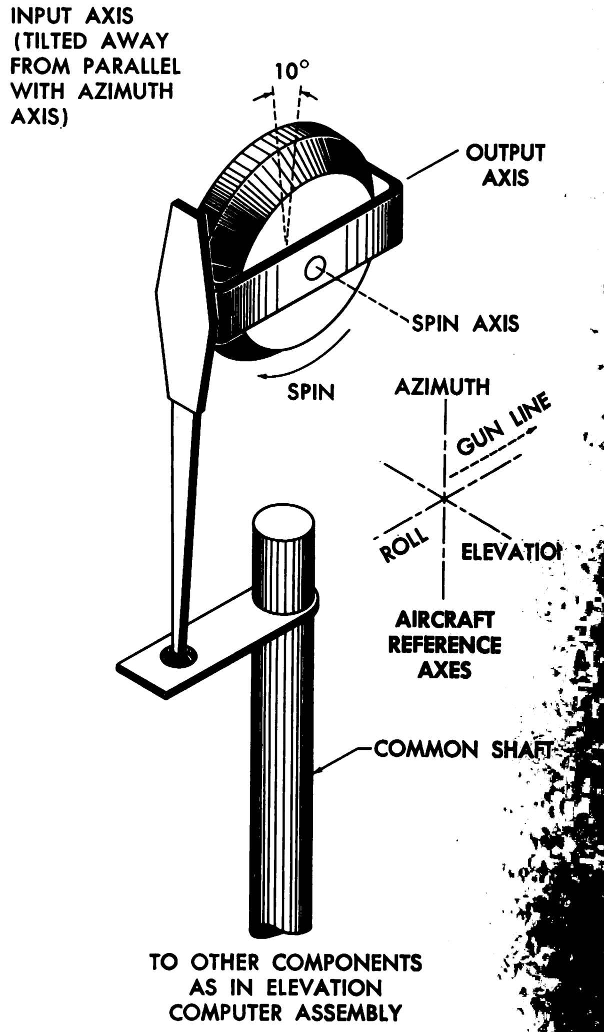


Computer Assembly (Elevation)

precesses away from you as you look at the diagram, the shaft will rotate in a clockwise direction. Any precession of the gyro will produce a corresponding shaft rotation with respect to both direction and amount. Furthermore, when the gyro is in the position of zero correction, the shaft assumes a corresponding position.

Although the gyro is fitted with mechanical stops, these stops alone do not make it a rate gyro. Operation as a rate gyro is the result of causing the gyro to precess against elastic restraint provided by the stiffness motor, which we will discuss more fully a little later.

The azimuth (or deflection) computer assembly is identical to the elevation computer assembly except for the gyro itself. The difference between the two gyros is largely a matter of changing the direction of the spin axis. This is shown in the illustration. Note

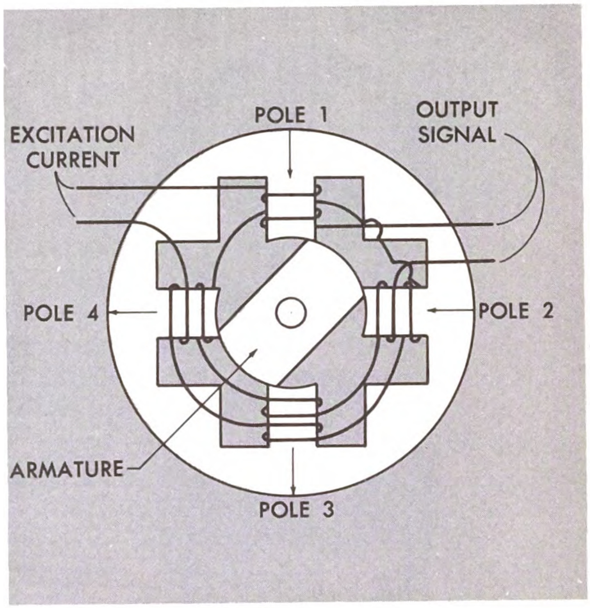


that the input axis is tilted 10° away from parallel with the azimuth axis. Since the input axis is tilted slightly away from parallel with the azimuth axis and slightly toward parallel with the roll axis, the amount of precession that results from any given rate of rotation about the azimuth axis is reduced slightly and a small amount of precession is produced by rotation about the roll axis. Making the azimuth gyro respond to a limited degree to rotation of the aircraft about its roll axis transfers a small portion of the elevation correction into the azimuth channel.

Note that only the elevation gyro is fitted with a gravity drop accelerometer, which is merely a small offset weight attached to the gyro gimbal. This weight tends to rotate the gyro about its output axis and, therefore, like precessional movement, will displace the common shaft from its zero correction position. In this manner, the gravity drop correction is introduced into the system. During the maneuvering of the aircraft, positive or negative *g* forces acting on the weight will increase or decrease its effect on the elevation gyro. Thus, the velocity jump correction is introduced by increasing or decreasing the gravity drop correction.

Pickoff. The pickoff is a device used to deliver an electrical signal which indicates the position of its armature. It consists of an armature surrounded by a four-pole magnetic structure as shown in the illustration. Each of the four poles has a coil consisting of two separate windings. These windings are connected into two series groups, each group consisting of one winding on each pole. The pickoff armature is coupled to the gyro through the shaft. Therefore, the position of the armature varies with the direction and amount of gyro precession.

A pickoff functions as a variable transformer. The four series-connected primary windings are connected to a low-voltage AC source. The resulting excitation current produces magnetic fields about the poles. The instantaneous polarities of these fields will



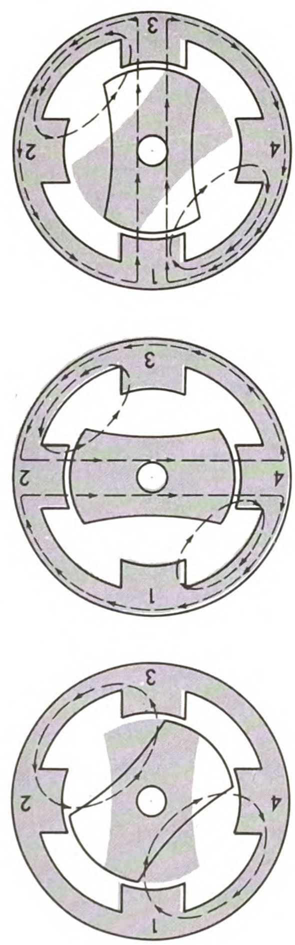
be as indicated by the arrows during one input half-cycle and just the opposite during the other half-cycle.

The four series-connected secondary windings are arranged so that the polarity of the voltages induced in coils *1* and *3* is always opposite that induced in coils *2* and *4*. Hence, the output voltage at any given time is equal to the difference between the sum of the voltages induced in coils *1* and *3* and the sum of the voltages induced in coils *2* and *4*. Further, the polarity of the output will be the same as that of the coils which produce the greater sum. This relationship can be expressed as follows:

$$E\ output = (E_1 + E_3) - (E_2 + E_4)$$

From this formula we can see that when the two sums are equal, the output voltage is zero.

The principles of operation involved in pickoffs can be seen by examining the magnetic circuit diagrams shown. The armature is just a piece of iron with no winding of any kind on it. When rotated, it varies the size of the air gaps between it and the opposing pairs of poles. In the upper diagram, the armature is shown at its neutral position. The same air gap exists between it and



Pickoff—Magnetic Circuits

each of the four poles. Therefore, the flux density of the two magnetic circuits indicated by the looped arrows will be equal, and identical voltages will be induced in all four secondary windings. Since the two opposing voltage sums are equal, the output voltage is zero. The mechanical linkage between the gyro and the pickoff is such that, when the gyro is in the position of zero correction, the pickoff armature is at its neutral position.

In the middle diagram, the armature is shown rotated clockwise into alignment with poles *2* and *4*. Note that the air gaps at poles *2* and *4* have been decreased. Because of the increased flux density through poles *2* and *4* and the decreased flux density through poles *1* and *3*, a greater voltage will be induced in coils *2* and *4* than in coils *1* and *3*. The result is an output voltage whose polarity is the same as that induced in coils *2* and *4*.

In the lower diagram, the armature has been rotated counterclockwise into alignment with poles *1* and *3*. Now the greater flux density passes through poles *1* and *3*, and the greater voltage is induced in coils *1* and *3*. Therefore, the polarity of the output voltage is the opposite to what it was in the previous case.

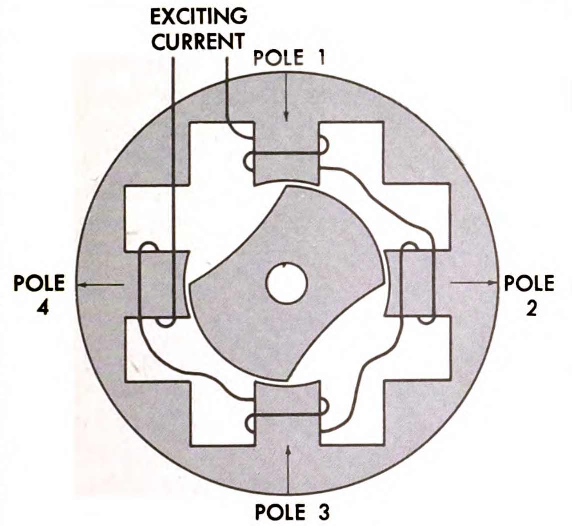
Hence, we see that, when the pickoff armature is rotated in one direction away from its neutral position, the output voltage assumes a certain polarity. When the armature is rotated in the opposite direction away from its neutral position, the output voltage assumes the opposite polarity. Therefore, the polarity of the output voltage is indicative of the direction of gyroscopic precession.

Actually the amount of pickoff armature rotation is limited to approximately 10°. This limitation, together with other design features, causes the output voltage to vary directly proportionally with the angular rotation of the armature away from its neutral position.

Summarizing the operation of a pickoff in conjunction with a gyro, we observe that it is used as the input measuring device of the servo loop. It translates the precessional movement of the gyro into an electrical signal which has two important characteristics: (1) the amplitude of the signal is proportional to the amount of gyro precession away from the position of zero correction, and (2) the polarity of the signal indicates the direction of precession away from the position of zero correction. This signal is then transmitted to the amplifier of the servo loop.

Stiffness Motor. The stiffness motor consists of an armature surrounded by a four-pole magnetic structure as shown in the diagram. Four windings, one on each pole, are connected in series. The directions of the

individual windings are such as to produce magnetic fields whose polarities are indicated by the arrows in the illustration. The armature is mounted on the same shaft as the pickoff armature.



The stiffness motor is used to control the amount of gyro precession by applying elastic restraint and also to provide a means of electrically caging the gyro. Current flowing through the windings of the stiffness motor tends to hold the armature at its neutral position which, incidentally, coincides with the neutral position of the pickoff armature and the gyro position of zero correction. The more current that flows through the windings, the greater the holding force or elastic restraint. Because of the mechanical linkage between the gyro and the stiffness motor armature, the gyro in precessing must rotate the stiffness motor armature away from its neutral position. Actually, the stiffness motor performs the same function as the springs in the basic rate gyro previously discussed.

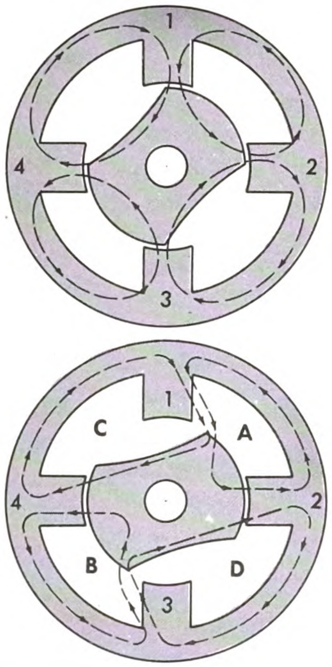
If the current through the stiffness motor windings is small, the elastic restraint is also small, and the gyro is comparatively free to precess. If the current increases, the elastic restraint also increases. This makes the gyro precession less for the same force acting upon the gyro. For electrical caging, maximum current is caused to pass through the windings, making the elastic restraint great enough to overcome any precessional force in the gyro. Hence, electrical caging is a means of zeroing the gyro. An important but perhaps less obvious function of the stiffness motor is to speed up the return of the gyro toward its zero correction position when the forces acting on the gyro diminish.

The stiffness motor makes use of the following magnetic principle: if a magnetic object is within a magnetic field and is free to move, the object will position itself so as to reduce the magnetic circuit through itself to the minimum possible length. Moving the object from this position will lengthen the flux lines, causing force to be exerted in opposition to the movement. The lengthening of the flux lines is comparable to the stretching of a rubber band.

In a stiffness motor, four magnetic circuits are formed through the armature and magnetic structure. These circuits are indicated by the closed loops of arrows in the following illustration. In the upper diagram the armature is shown at its neutral position because all four of the magnetic circuits are as short as they can possibly be.

Now let us see what happens when the armature is rotated away from its neutral position. In the lower diagram the armature has moved in a clockwise direction. Because of the increased air gaps at poles *1* and *3*, and because flux lines prefer a metallic path, all four magnetic circuits are lengthened at *A* and *B*. Had the armature been turned in the opposite direction away from its neutral position, a similar lengthening of the flux lines would have taken place at *C* and *D*. In either case the lengthened magnetic circuits will tend to pull the armature back to its neutral position.

The amount of elastic restraint produced by the stiffness motor is determined by two factors. It is proportional to the square of



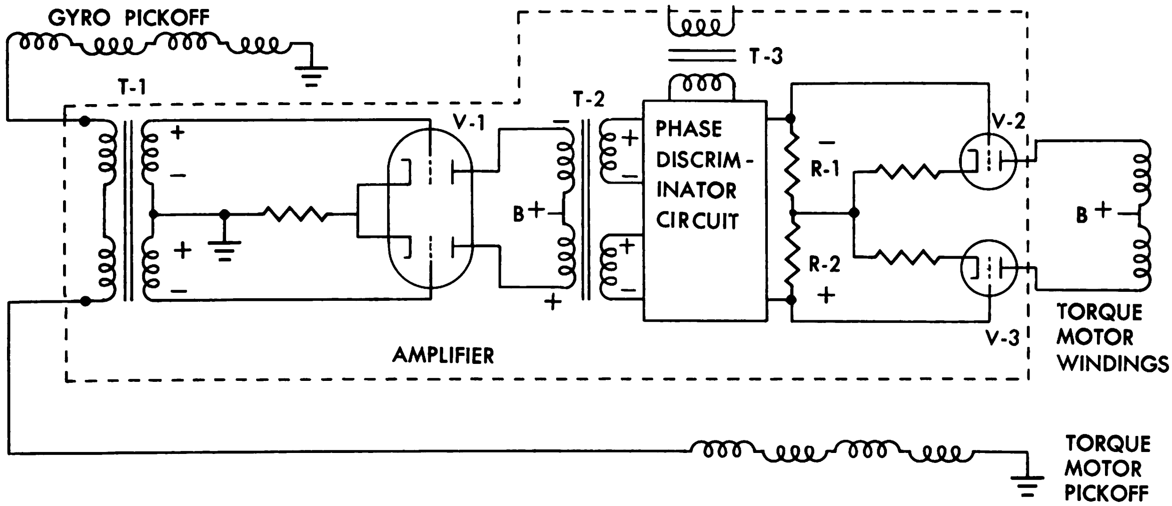
the current passing through the windings, and it is practically directly proportional to the angular displacement of the armature away from its neutral position. This means that a given value of current establishes the basic amount of elastic restraint that tends to oppose the movement of the armature away from its neutral position and that this amount of elastic restraint increases directly proportionally to the angular displacement of the armature from its neutral position.

The current which excites the stiffness motor windings is provided by the sensitivity amplifier. The amplitude of this current is determined by the range, air density, and airspeed inputs to the computing system. At any given instant the value of the current is inversely proportional to the projectile time of flight to the future target position at that time. Because of the mechanical linkage between the gyro and the stiffness motor, all the factors that are inputs to the computing system affect the amount of gyro precession. Therefore, the amount of gyro precession is the amount of correction worked out by the computing system.

Viscous Damper. The viscous damper consists of a disc-shaped paddle and a damper well. The paddle is positioned within the damper well, and the small space between the two is filled with damping fluid. This fluid, being both thick and sticky, adheres to the surfaces of the paddle and well. In doing so, it tends to prevent relative motion by creating friction between the two. The effect is practically the same as if the surface of the paddle were rubbing against the surface of the well, but the use of a fluid eliminates mechanical wear. Any force which tries to rotate the paddle with respect to the well will have part of its energy used up in overcoming this friction. The extent of the resistance offered by the viscous damper depends on two factors: the relative speed between the paddle and the well, and the viscosity (thickness) of the fluid.

In the computer assembly, one of the purposes of the viscous damper is to stabilize the gyro by smoothing out jerky precessional movement. Disregarding for a moment the effect of the stiffness motor, the mechanical coupling between the gyro and the viscous damper causes the amount of rotation of the common shaft to depend on two factors: the amount of precessional force developed by the gyro, and the length of time that force is exerted. Brief, sudden changes in the precessional force of the gyro will therefore be largely absorbed in overcoming the friction of the viscous damper.

Another function of the viscous damper is to provide for the calibration of the computer. Adjustable thermostatically controlled heaters are used in conjunction with the viscous dampers in the computer assembly. Since the friction of the damper depends on the viscosity of the fluid and the viscosity depends on the temperature, changing the temperature of the fluid will cause the viscous damper to absorb more or less of the precessional force produced by the associated gyro. Thus, for a given set of inputs to the computer, we can adjust for the proper *total* output by setting the thermostat of the damper heater.



THE AMPLIFIER. A simplified diagram of the elevation or azimuth amplifier together with its input and output connections is shown.

The signal windings of the gyro pickoff and the torque motor pickoff are connected to form a series circuit with the primary of input transformer *T-1*, which acts as the differential in the basic servo loop previously discussed. Therefore, the effective voltage applied to the primary of *T-1* is always equal to the algebraic sum of the voltages induced in the signal windings of the two pickoffs. Whenever the armatures of the two pickoffs are in the position of correspondence, the algebraic sum is zero.

The basic situation in following through the operation of the amplifier is one in which the armatures of both pickoffs are in the neutral position. This means that the gyro is not precessed, and that the prediction mirror is positioning the reticle image for zero correction.

The input voltage to the primary of *T-1* is zero because the output of each pickoff is zero. Hence, no voltage is induced in the secondaries of *T-1*. The only signal applied to the grids of tube *V-1* is the bias voltage dropped across its cathode resistor. Since both triodes of *V-1* receive the same grid bias, they will conduct equal plate current through the respective halves of the primary of coupling transformer *T-2*.

The plate currents of the two triodes flow in opposite directions through the halves of the primary of *T-2*. Being equal and opposite, they cancel each other. Therefore, the effective input to *T-2* is also zero, and the voltage induced in the secondaries of *T-2* is also zero. The operation of the phase discriminating circuit is such that, when the output of *T-2* is zero, no current flows through load resistors *R-1* and *R-2*, and the voltage drop across *R-1* and *R-2* in series is also zero.

With no voltage dropped across *R-1* and *R-2*, the only signal applied to the grids of tubes *V-2* and *V-3* is the bias dropped across their respective cathode resistors. Since these resistors have the same value, the same grid bias is applied to the two tubes. Therefore, tubes *V-2* and *V-3* conduct equal plate currents. Equal plate currents flowing through the opposing windings of the torque motor cause its armature to be held at its neutral position, which corresponds to the neutral position of the armature of the torque motor pickoff.

Note that nothing has happened to disturb the initial relationship between the armatures of the two pickoffs. The error signal appearing at the primary of *T-1* remains

zero. The torque motor maintains its armature at the neutral position. The entire servo loop may be considered "at rest" under these conditions.

Now let us see what happens when the gyro precesses as it does when a target is tracked. The gyro precesses through an angle whose magnitude and direction are determined by the various inputs to the computer assembly. This rotates the gyro pickoff armature through a corresponding angle, and causes a proportional voltage to be induced in the signal windings of the gyro pickoff. This signal voltage is initially the error signal, because the torque motor pickoff armature has not yet moved from its neutral position. The phase of the error signal depends on the direction the gyro pickoff armature was displaced from its neutral position. In one direction, the error signal will be in phase with the excitation; in the other direction, it will be 180° out of phase.

Assume that the phase of the error signal is such as to induce voltages which make positive the upper ends of the two secondaries of *T-1*. The grid of the upper triode of *V-1* becomes more positive; that of the lower triode becomes more negative. The plate current of the upper triode increases, while that of the lower triode decreases. More current flows through the upper half of the primary of *T-2* than through the lower half. The direction of the net effective current produces a negative polarity at the upper end of the primary of *T-2* and a positive polarity at the lower end. The direction of the effective current is such as to induce voltages of the indicated polarities in the secondaries of *T-2*.

The voltages induced in the secondaries of *T-2* are applied to the phase discriminator circuit together with the reference voltage from transformer *T-3*, which is always in phase with the excitation of the pickoffs. With the input from *T-2* as indicated, the phase discriminator circuit causes current flow downward through resistors *R-1* and *R-2*. This produces a voltage drop across the two resistors that is negative at the upper end of *R-1* and positive at the lower end of *R-2*. The amplitude of the voltage drop is proportional to the amplitude of the error signal that initiated the sequence of operation.

The grid of tube *V-2* is made more negative by the voltage drop, and at the same time the grid of tube *V-3* is made more positive. Tube *V-2*, therefore, conducts less current through the upper pair of torque motor windings than tube *V-3* conducts through the lower pair. Since one pair of windings exerts a greater torque on the motor armature than the other pair, the torque motor armature is displaced in a certain direction away from its neutral position. This displacement is transmitted mechanically to the prediction mirror and to the torque motor pickoff. The movement of the mirror displaces the reticle image on the windshield or combining glass, thereby beginning to generate the correction angle between the sight line and the gun line.

Movement of the torque motor pickoff armature away from its neutral position begins at the same time as the displacement of the prediction mirror. This induces in the signal windings of this pickoff a voltage which, when applied to the primary of *T-1*, opposes the signal from the gyro pickoff. Since the error signal at the primary of *T-1* is the algebraic sum of the two signal voltages, the error signal approaches zero as the two signals approach equality. When the error signal actually reaches zero, the prediction mirror has been displaced through the angle determined by gyro precession, and the proper correction angle has been established between the sight line and the gun line.

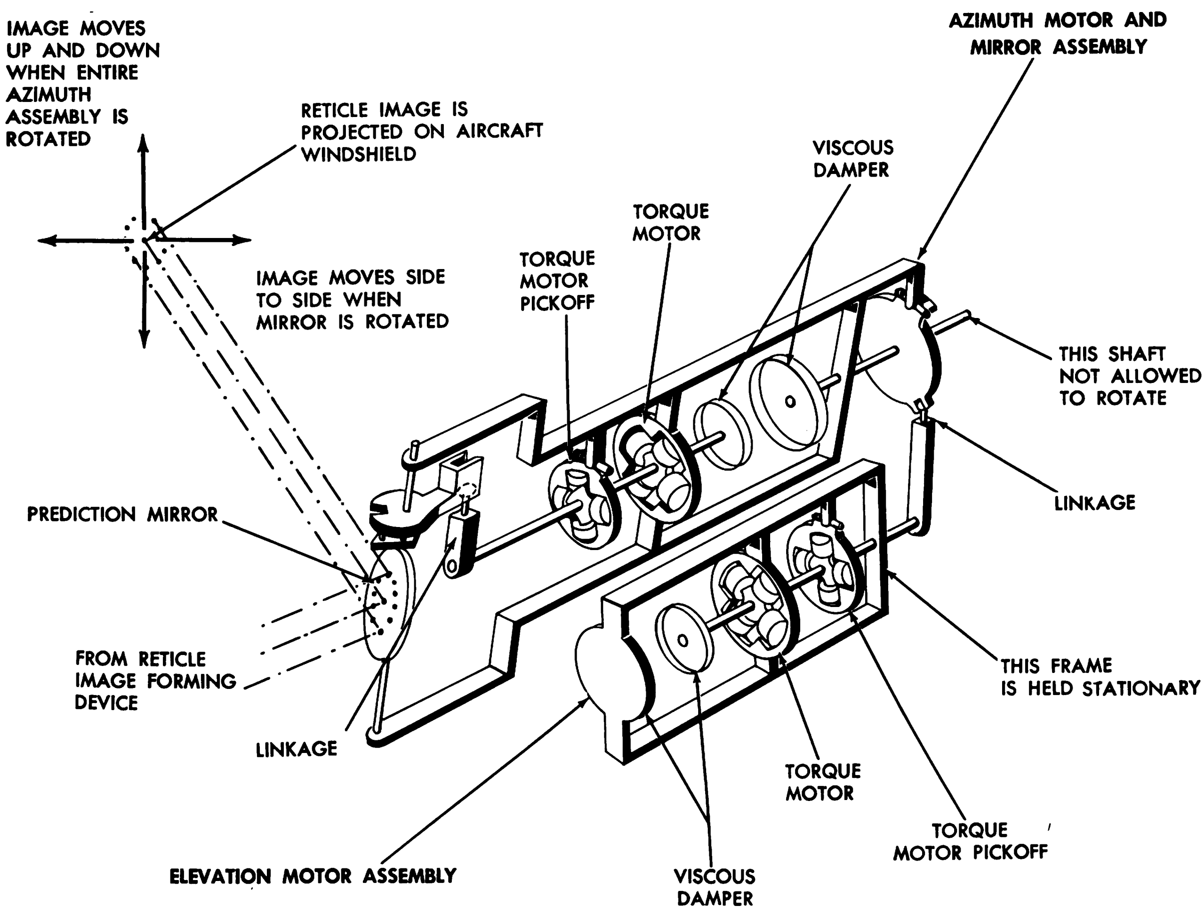
When the error signal becomes zero, the two pairs of windings of the torque motor again receive equal currents. This condition tends to cause the torque motor armature to return to its neutral position. However, the torque motor armature cannot move without moving the torque motor pickoff. Therefore, if the torque motor armature tries to return to neutral, the torque motor pickoff armature will be displaced from its

zero-error-signal relationship with the gyro pickoff armature. This would create an error signal whose effect would be to drive the torque motor armature away from its neutral position and the torque motor pickoff armature toward the zero-error-signal position.

If the gyro had precessed away from its neutral position in the direction opposite to what we have considered, the phase of the gyro pickoff output would shift 180°. This would reverse all the polarities shown in the previous illustration, except that of reference transformer *T-3*. The final outcome would be that the greater current would flow through the upper pair of torque motor windings. Opposite torque would be exerted on the torque motor armature, driving the prediction mirror and the reticle image in the opposite direction.

Mirror Drive Assembly. The mirror drive assembly is composed of two subassemblies: the azimuth mirror and motor assembly, and the elevation motor assembly. The physical relationship between the components of each subassembly and between the two subassemblies is shown in the simplified mechanical schematic diagram.

Azimuth Mirror and Motor Assembly. This assembly consists of torque motor, pickoff, viscous damper, and two mechanical linkages. One linkage couples the torque motor to the prediction mirror, which is also part of this assembly; the other linkage couples the two subassemblies.



Mirror Drive Assembly

The armatures of the torque motor and pickoff, the paddle of the viscous damper, and the lever of the linkage to the mirror are all mounted on the same shaft. Rotation of this shaft is produced by operation of the torque motor in response to the output of the azimuth amplifier. The only opposition to the rotation of the shaft is the friction in the viscous damper. The shaft rotation is translated by the mechanical linkage into rotation of the mirror about its own axis. This causes the reticle image to move laterally (side-to-side) across the windshield or combining glass. At the same time the pickoff transmits to the azimuth amplifier a signal indicative of shaft position. Since the mirror is positioned by the shaft, the signal is also indicative of mirror position.

The entire azimuth mirror and motor assembly is mounted in the sight head so that it can rotate about an axis which coincides with that of the shaft. Rotation of the entire assembly about this axis changes the tilt angle of the mirror relative to the windshield. This causes the reticle image to move vertically (up and down) on the windshield or combining glass.

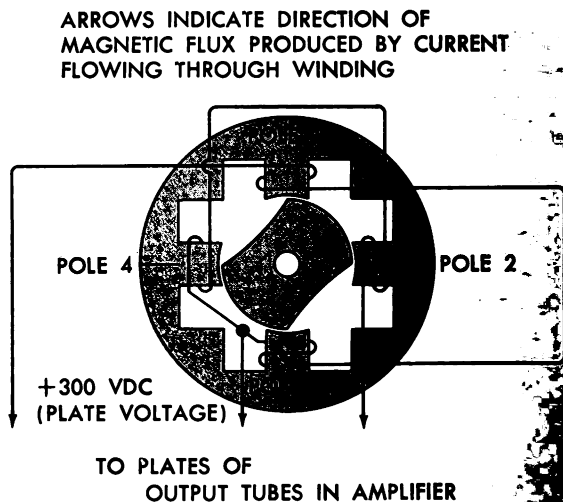
Elevation Motor Assembly. This assembly consists of a torque motor, pickoff, viscous damper, and part of the linkage to the azimuth mirror and motor assembly. The armatures of the torque motor and pickoff, the paddle of the viscous damper, and the lever of the linkage are mounted on the same shaft. The shaft is the only part of the assembly capable of movement within the sight head.

Operation of the torque motor in response to the output of the elevation amplifier produces rotation of the shaft. The only opposition to shaft rotation is friction in the viscous damper. Rotation of the shaft not only positions the pickoff armature, but rotates the lever of the linkage. This causes the entire azimuth mirror and motor assembly to rotate about its axis.

Viscous Dampers. The purpose of the viscous damper used in conjunction with the torque motor is to minimize overshooting or hunting which might result from momentum built in the load of the torque motor. The heaters of the viscous dampers used in mirror drive assemblies have no calibration adjustment. They are operated at a thermostatically controlled constant temperature.

Pickoffs. The pickoffs used in the mirror drive assembly are identical to those used in the computer assemblies.

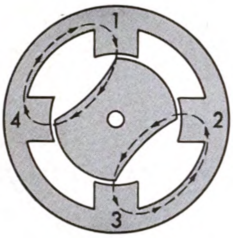
Torque Motors. A torque motor consists of an armature surrounded by a four-pole magnetic structure as shown in the illustration. Four identical windings, one on each pole, are connected so that each pair of windings on opposite poles constitutes one half of the load of the azimuth or elevation amplifier, as the case may be.



Unlike ordinary motors, the torque motor does not produce continuous rotation. Its armature is limited to movement through a small angle to either side of the neutral position. The magnitude and direction of armature rotation are determined by the output of the amplifier as already described.

When the error signal is zero, the amplifier output current flows equally through all four coils. This produces magnetic fields of equal intensity through the four poles. The direction of the windings is such that mag-

netic circuits form as shown by the closed loops of arrows in the next illustration.

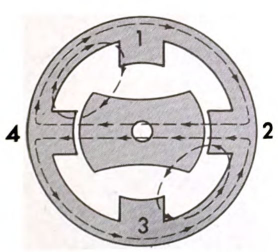


Torque Motor—Neutral Positions

The two magnetic circuits, each having the same flux density, position the armature at the place where they exert equal and opposite torques or turning forces. The neutral position, therefore, is that in which the two magnetic circuits are of equal length.

When the error signal is not zero, the currents through the pairs of torque motor windings are unequal. As the error signal increases, the current through one pair of torque motor windings increases and that through the other pair decreases. The phase or polarity of the error signal determines which pair of windings receives greater or lesser current.

Let us assume that the phase of the error signal is such that the current through windings *2* and *4* becomes greater than that through windings *1* and *3* in the related illustration. The number of flux lines passing through poles *2* and *4* is now greater than that through poles *1* and *3*. The torque tending to aline the armature with poles *2* and *4* becomes greater than the torque toward alinement with poles *1* and *3*. Therefore, the armature rotates clockwise through an angle proportional to the difference between the two torques.



Torque Motor—Clockwise Movement

If the error signal were of the opposite phase, windings *1* and *3* would receive more current than windings *2* and *4*. The armature would then tend to aline itself with poles *1* and *3*. In this case, the armature would move counterclockwise from its neutral position through an angle proportional to the difference between the two torques.

DETERMINATION OF STIFFNESS MOTOR CURRENT. We have already seen that the prediction angle evolved by the computer is produced by motion of the prediction mirror in response to precession of the azimuth and elevation gyros. In the course of our discussion of this operation, we covered the stiffness motor from the viewpoint of how it works and how it affects gyro precession. However, to prevent confusing our operational analysis of the azimuth and elevation servo loops with a sidetrip through three other servo loops, we took only a cursory glance at the source of stiffness motor current and merely observed that the value of the current is inversely proportional to time of flight.

A better understanding of how all the factors that affect the prediction angle are integrated into the precession of the gyros demands that we consider in more detail the source of stiffness motor current and how its value is determined. Because of the numerous components involved, this will be no easy matter. As a result, our discussion will be centered on block diagrams which show the mechanical and electrical relationships between the various components in the most simple manner possible.

Before getting into our detailed discussion, let us briefly preview the tie-in of the three interrelated servo loops that we deliberately avoided earlier in this chapter. The range signal provided either by the radar set or the manual range control is the

input to a servo loop whose principal component is the range servo unit. This servo loop produces three outputs which vary according to the range signal. One output is supplied as an input to a servo loop whose principal components are the range amplifier and the reticle forming device. The remaining two outputs are applied as inputs to a servo loop whose principal components are the sensitivity amplifier and the stiffness motors.

Range Servo Unit. The range servo unit consists of five packaged plug-in subunits—adapter, modulator, amplifier, mechanism, and power supply—mounted on either one or two chassis bases. (In some types of aircraft two smaller assemblies are more conveniently installed than one large assembly.) When two chassis bases are used, the mechanism is mounted on one, and the remaining four subunits on the other.

The range servo unit is not merely an auxiliary device. It performs a number of functions, some of which make it an integral part of the computing system. Rather than list the various functions at this time, let us present them as we come to them in the following discussion. But in doing so, we must keep in mind that, with respect to the computing system, the principal function of the range servo unit is to convert a signal indicative of range into signals indicative of the *effect of range* on the flight of the projectile.

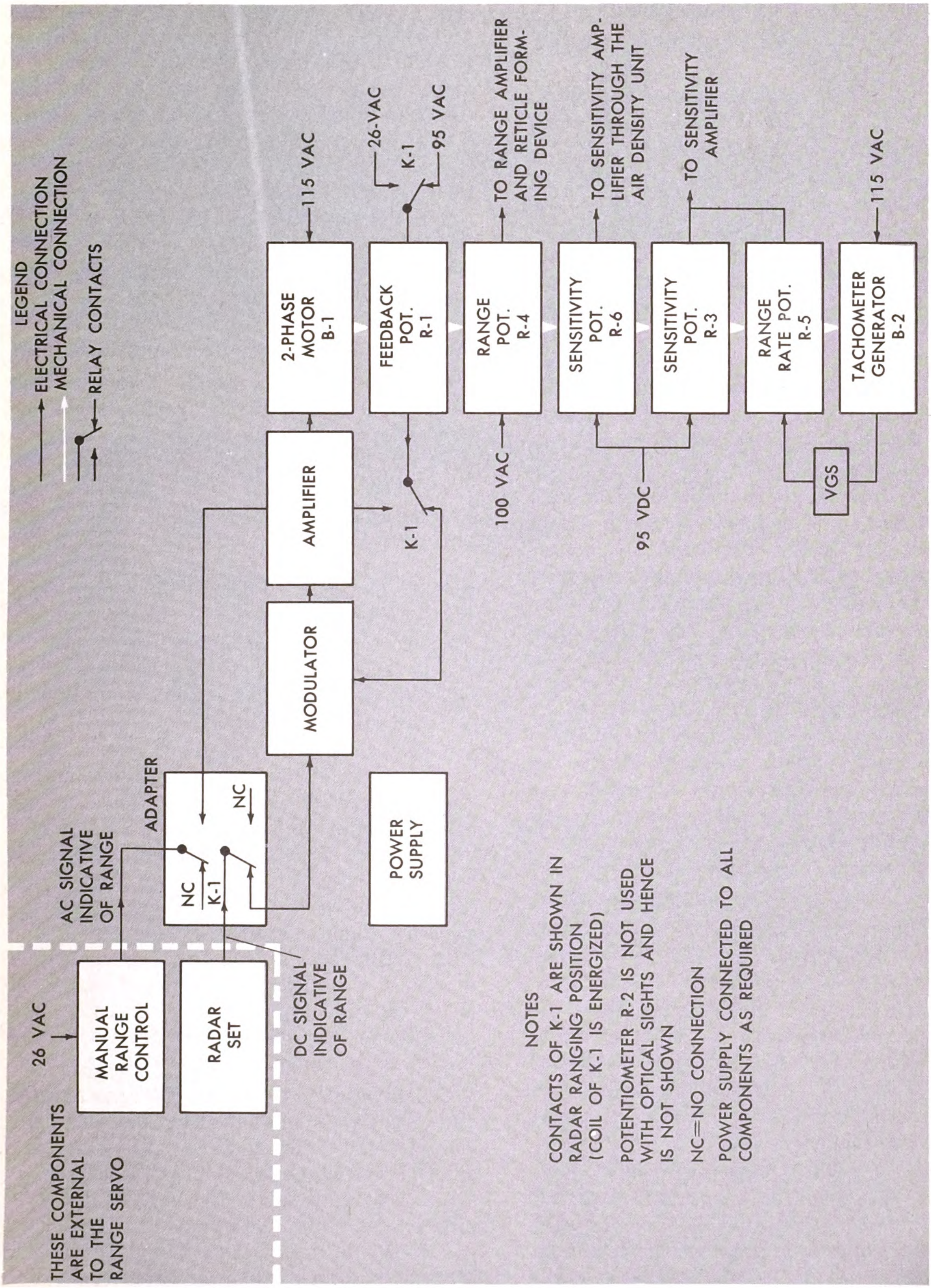
The mechanical and electrical relationships between the major components of the range servo unit are shown in the block diagram presented. Four of the subassemblies—adapter, modulator, amplifier, and power supply—are represented by single blocks. All of the remaining blocks except those indicating external signal sources are components of the mechanism.

In our discussion of the servo loop of the range servo unit we can for the moment eliminate all components of the mechanism which are input measuring devices for the other two servo loops. Included in this category are all components mechanically connected to motor *B-1* except feedback potentiometer *R-1*. This leaves us with a servo loop in which the radar set or the manual range control is the input measuring device, the modulator and/or the input circuit of the amplifier is the differential, motor *B-1* is the output, and feedback potentiometer *R-1* is the output measuring device. The operation of this servo loop will be analyzed with respect to radar ranging first and then manual ranging.

In *radar ranging*, relay *K-1* in the adapter is energized, and its movable contacts establish the connections shown in the block diagram. (Actually all four sets of contacts are located within the adapter, but to simplify the diagram some of them have been shown outside.) The contacts of relay *K-1* disconnect the output of the manual range control from the range servo unit, connect the modulator to the radar set output voltage whose amplitude is indicative of range, connect the output of feedback potentiometer *R-1* to the modulator, and connect *R-1* to a regulated DC voltage source.

The operational characteristics of the amplifier are such that an AC error signal is required as an input. But during radar ranging both the input and output signals are DC. Hence, the function of the modulator is to convert these two signals into an AC error signal. Both signals are applied to a chopper in the modulator. The chopper produces a 400-cycle square wave, whose amplitude is equal to the difference between the amplitudes of the two signals. When the two signals are equal, the amplitude of the square wave is zero. Electronic circuits in the modulator convert the square wave to a 400-cycle, sine-wave AC, whose amplitude is still proportional to the difference between the two signals. This 400-cycle AC is the error signal that is fed to the amplifier. The amplifier builds up the error signal and produces a proportional AC output.

The phase of the amplifier output depends on whether the range signal is increasing or decreasing. If the range signal is increasing, we will have a certain output phase



Range Servo Unit

Then, if the range signal changes from increasing to decreasing, the amplifier output phase will shift 180°.

The amplifier output goes to the control winding of motor *B-1*, the other winding being connected to fixed-phase, 115-volt AC. One phase of amplifier output will make the motor run in one direction; the other phase will make it run in the other direction. Therefore, the motor will run in a certain direction when the range is increasing and in the other direction when range is decreasing. The motor will run until the feedback or output signal is equal to the range signal. Hence, all devices mechanically coupled to the motor are positioned according to range as indicated by the radar output voltage.

This brings us to an important function of the adapter. It connects feedback potentiometer *R-1* to the regulated DC source through an adjustable resistance network which is not shown in the block diagram. This network enables the feedback voltage of *R-1* to vary according to the range indication of motor *B-1* in exactly the same manner that the radar output voltage varies according to actual range. Otherwise, the servo loop will not function properly. Hence, the function of the adapter is to match the range servo unit to a particular radar set. By changing adapter subunits, we could accommodate some other type of radar set.

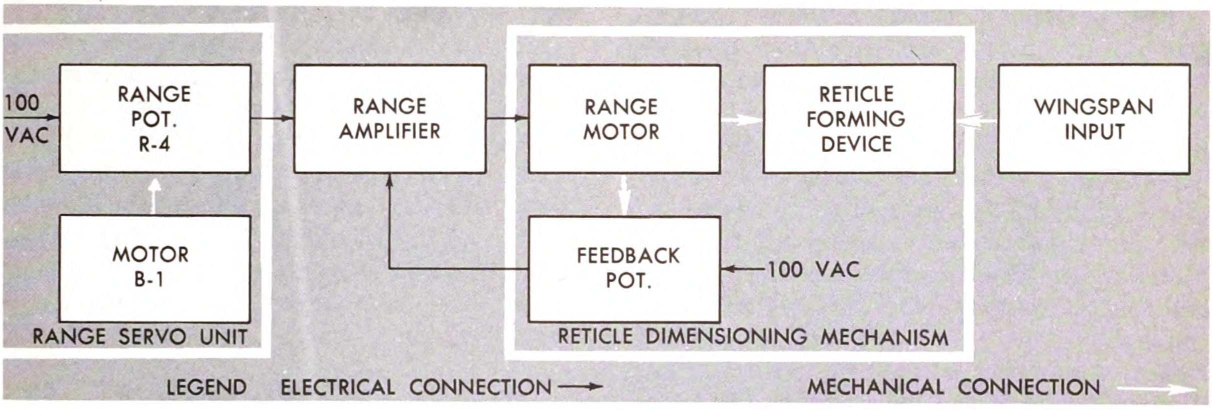
In *manual ranging,* relay *K-1* is deenergized, and its movable contacts go over to the opposite stationary contacts. The manual range control is connected to the amplifier, the radar output is disconnected from the modulator, the output of *R-1* is connected to the amplifier, and the excitation of *R-1* is switched to an AC source of the same amplitude as that applied to the manual range control but of opposite phase.

When the pilot rotates the manual range control, he sends to the amplifier an AC signal whose amplitude varies with the amount of rotation of the control. The output of the amplifier drives the motor as in the case of radar ranging. The motor positions the wiper of *R-1*, which now sends back an AC signal indicative of motor position. The motor continues to run until the feedback signal is equal to the manual range signal. Being of opposite phase, the two signals cancel each other when equal, producing a zero signal. When the two signals are unequal, the amplitude of the error signal is the difference between them, and the phase of the error signal is the same as that of whichever signal is the greater. (Since no conversion to an AC error signal is required, the modulator is not used in manual ranging.)

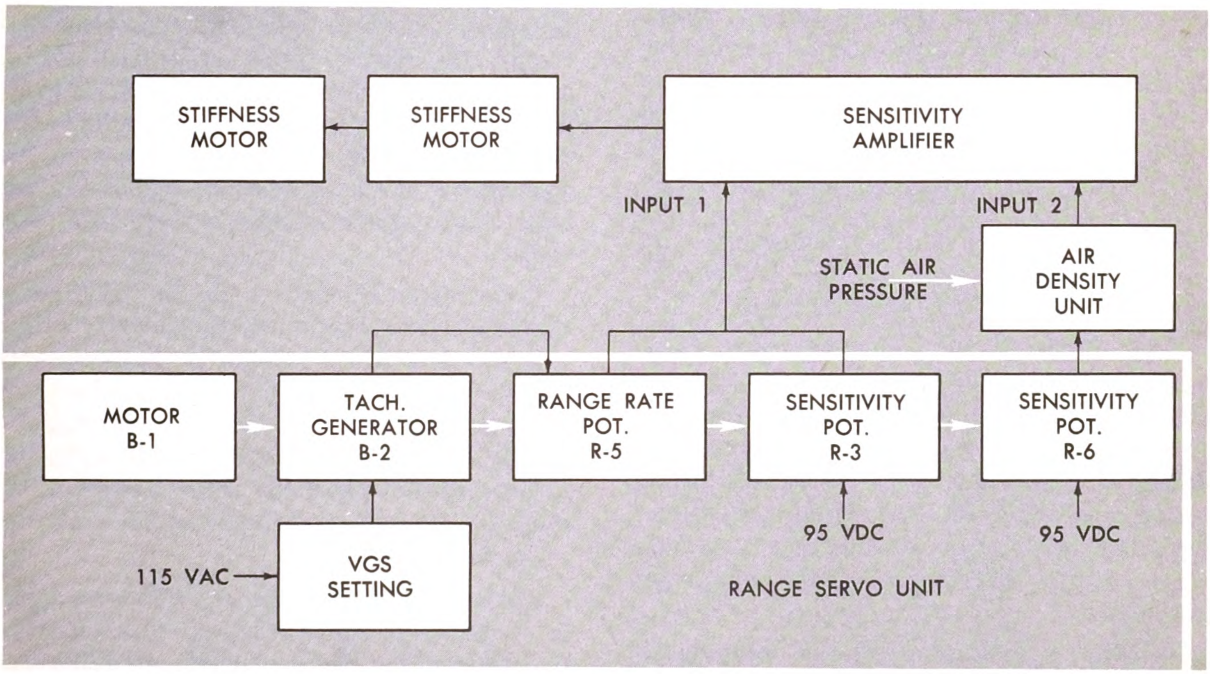
Servo Loop of the Range Amplifier. The mechanical and electrical relationships between the components of this servo loop are shown in the accompanying block diagram. The input measuring device of this servo loop is range potentiometer *R-4* in the mechanism of the range servo unit. The wiper of *R-4* is positioned by motor *B-1*, which in turn is driven in response to the range signal provided by either the radar set or the manual range control. Hence, the output of *R-4* is an AC voltage whose amplitude is indicative of range. This output is applied as an input to the range amplifier.

The operation of this servo loop is quite conventional. The output of the range amplifier energizes the range motor, which drives both the reticle forming device and the wiper of the feedback potentiometer. Since there is a direct relationship between the mechanical input to the reticle forming device and the electrical range signal that enters the range servo unit, the reticle forming device is driven to a position indicative of range. If the wingspan input is set correctly, the diameter of the reticle image circle will vary according to range.

Sensitivity Servo Loop. The mechanical and electrical relationships between the components of the sensitivity servo loop are shown in the next block diagram. In the mechanism of the range servo unit, motor *B-4* is geared to three potentiometers (*R-5*, *R-8*, and *R-6*) and to induction motor *B-2*, which is used as a tachometer generator.



Servo Loop of the Range Amplifier



Sensitivity Servo Loop

It has already been shown that motor *B-1* is driven in response to the range signal. Because of their mechanical coupling to this motor, the wipers of all three potentiometers are driven to a position indicative of range. The resistance element of each of these potentiometers is tapered to conform with a different function of range. (A function of range is a mathematical equation in which range is the independent variable. In the case of these potentiometers, the equations express the effects of range variation on projectile time of flight.) Hence, if we apply a voltage to a resistance element and vary the position of the wiper according to range, the output voltage measured at the wiper will vary according to some effect that range has on time of flight.

The operational characteristics of tachometer generator *B-2* are such that the amplitude of its output depends on the rate at which it is driven and the phase of its output depends on the direction in which it is driven. As a result, the tachometer generator delivers an AC signal that indicates the rate at which range is changing and the direction in which range is changing (increasing or decreasing). This signal, when combined with an indication of present range, enables the computing system to anticipate future target range under the existing conditions.

The excitation voltage of tachometer generator *B-2* is varied according to the VGS (velocity of the gun station) setting, which enters the speed of the interceptor as a factor in the calculation of time of flight. Up to now, optical computing systems have used either no VGS setting at all or one that is preset prior to takeoff. It is possible to take VGS as a continuously variable factor by using ram pressure and a transducer so that the excitation voltage of the tachometer generator varies with true airspeed.

The output of the tachometer generator provides the excitation voltage for the resistance element of range rate potentiometer *R-5*. The output of *R-5* is an AC signal indicative of the combined effects of range, range rate, and VGS. This signal is combined with the DC output of sensitivity potentiometer *R-3* to form input *1* of the sensitivity amplifier.

Sensitivity potentiometer *R-6* provides a DC output indicative of the third effect of range on time of flight. This signal, after being varied according to static air pressure by the air density unit, becomes input *2* to the sensitivity amplifier. Note that inputs *1* and *2* together incorporate all the factors that determine time of flight to the future target position: range rate, gun station velocity, air density, and three effects of present target range.

The sensitivity amplifier converts inputs *1* and *2* into an AC error signal, which is first amplified, then rectified, and finally applied to a DC amplifier consisting of several triodes in parallel. The output of the DC amplifier is a current whose amplitude is inversely proportional to the time of flight. This current passes through the windings of both stiffness motors connected in series. The resulting elastic restraint produced by the stiffness motors enters time of flight as a factor in determining the amount of precession by the gyros.

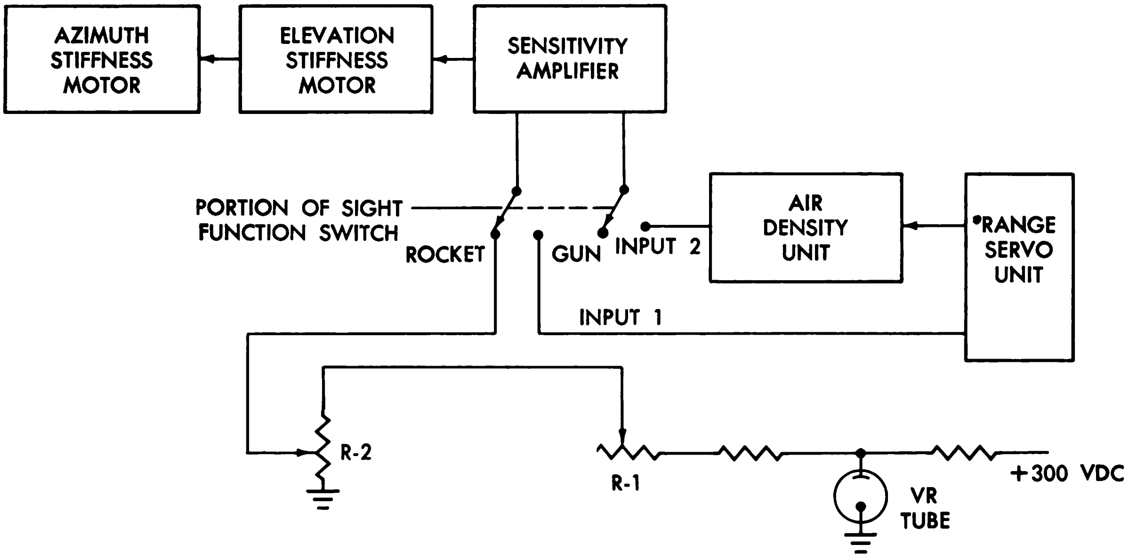
Special Considerations in Rocketfire and Bombing Functions

Optical computing sights generally include components which provide for air-to-ground rocketfire and dive bombing functions. Although the gunfire function is used exclusively in interceptor operations, the rocketfire and bombing functions are available for possible use. This is in accordance with the concept of interceptor capability for other operations presented in chapter 1 of this manual.

ROCKETFIRE FUNCTION. Special considerations must be given to the differences in time of flight and gravity drop between rockets and projectiles fired from machine guns or cannon. The nature of these considerations is such that we can anticipate changes in the sensitivity and elevation servo loops in setting up the computing system for the rocketfire function.

Rocket Sensitivity Circuit. When the sight is switched to the rocketfire function, the two inputs that the sensitivity amplifier receives from the range servo are disconnected as shown in the simplified diagram. Instead, the sensitivity amplifier receives a single input from the rocket sensitivity circuit.

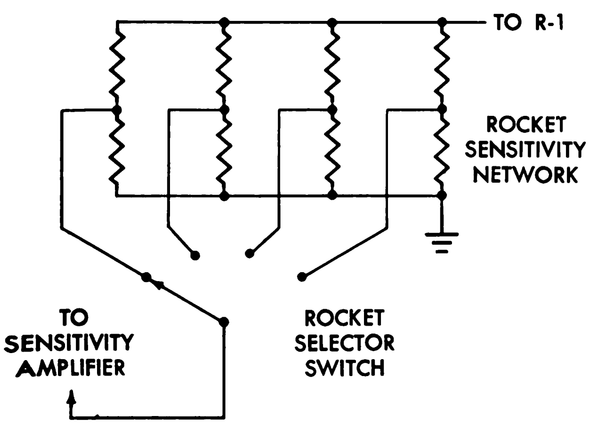
The basic circuit is quite simple. The voltage regulator tube in conjunction with the two dropping resistors on either side of it cause a stabilized DC voltage to be applied to the series combination of resistors *R-1* and *R-2*. Resistor *R-1* determines the voltage applied to resistor *R-2*. Resistor *R-2* acts as a voltage divider, the ratio of the re-



Rocket Sensitivity Circuit

sistance above the wiper to the resistance below the wiper determines the voltage that is applied as the single input to the sensitivity amplifier.

The setting of *R-2* depends solely on the type of rocket being fired. The output voltage of *R-2* is such as to cause the sensitivity amplifier to deliver to the stiffness motors an excitation current whose value corresponds to the time of flight of the particular type of rocket over its maximum effective range.



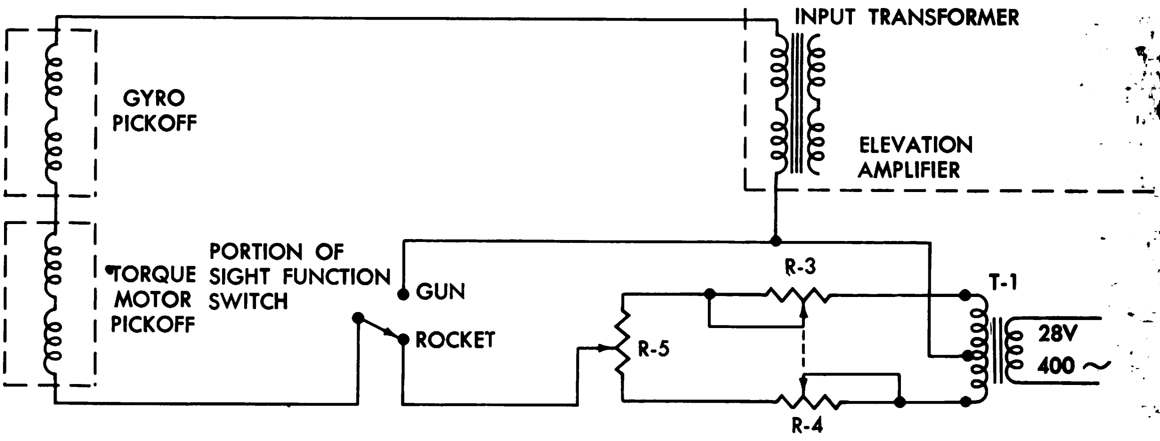
Variation of Rocket Sensitivity Circuit

Hence, the azimuth and elevation lead corrections are based on a fixed range input for each type of rocket.

In some sights a resistor network and selector switch are used in place of potentiometer *R-2* in the rocket sensitivity circuit just described. This arrangement is shown in the accompanying diagram. Here the basic operation is the same as when *R-2* alone was used. In this case, a pair of series resistors determines the input voltage to the sensitivity amplifier for each type of rocket. The resistance values of each series pair are selected to produce the proper value of stiffness current for each specific type of rocket.

Rocket Gravity Correction Circuit. This circuit is used to insert into the elevation servo loop a signal which changes the elevation correction angle according to the difference between the gravity drop of the particular type of rocket and the gravity drop of the projectile fired in the gunfire function. A simplified version of this circuit is shown in the next diagram.

During the gunfire function, the input circuit of the elevation amplifier consists of a series arrangement of gyro pickoff, torque



Rocket Gravity Correction Circuit

motor pickoff, and input transformer. When the system is switched to the rocketfire function, the circuit consisting of transformer *T-1* and potentiometers *R-3*, *R-4*, and *R-5* is connected in series with the three components just listed.

When the wiper of potentiometer *R-5* is at the electrical midpoint of the resistance, no voltage is inserted into the amplifier input circuit by the rocket gravity correction circuit. When the wiper is moved in one direction away from the midpoint, a voltage proportional to the displacement and in phase with the gyro pickoff output is inserted into the circuit. When the wiper is moved in the opposite direction away from the midpoint, the inserted voltage is out-of-phase with the gyro pickoff output.

If the inserted voltage is in-phase, the inserted voltage is added to the gyro pickoff output, creating a larger error signal. This means that the prediction mirror must move through a greater angle to make the torque motor pickoff output large enough to cancel the increased input signal. Hence, the elevation correction angle is increased in this case.

If, on the other hand, the inserted voltage is out-of-phase, the total input signal is diminished because the inserted voltage will partly cancel the gyro pickoff output. Now the servo loop operates to reduce the elevation correction angle because a smaller torque motor pickoff output is capable of canceling the input signal.

The setting of potentiometer *R-5* is determined by two factors: (1) the type of rocket, and (2) the dive angle at which the rocket is released. Potentiometers *R-3* and *R-4* provide the rocket boresighting adjustment in the harmonization procedure.

BOMBING FUNCTION. When an optical sight is set up for the bombing functions, certain mechanical and electrical changes occur in the elevation and azimuth servo loops. First of all, a bombing accelerometer is mechanically coupled to the elevation gyro. Basically, the bombing accelerometer is an off-center weight that causes the gyro to tilt as it does when it precesses. The weight is sufficient to force the gyro down as far as an adjustable mechanical stop will permit. A typical bombing accelerometer will be described later in this chapter.

Since time of flight correction is not necessary in the bombing function, the outputs of the sensitivity and range servo loops are disconnected from the stiffness motors and range motor respectively. Obviously then, both gyros operate with zero elastic restraint except during electrical caging.

The elevation torque motor is disconnected from the elevation amplifier. One half of the motor windings are energized from a con-

stant power source. The result is that the motor drives the prediction mirror down against its mechanical stop, which causes the reticle image to be depressed about 10°. Remember that the bombing accelerometer drove the elevation gyro all the way down, too. The error signal to the elevation amplifier is zero because the gyro and torque motor pickoff armatures have been driven to identical positions.

In place of the elevation torque motor, the coil of a differential relay is connected to the output of the elevation amplifier. The differential relay does not operate unless there is amplifier output. For the present, the error signal is zero, and so is the amplifier output. To make the differential relay operate, an error signal must be created, and this can be done only by raising the elevation gyro off its mechanical stop.

The depression of the reticle image makes the aircraft follow a downward spiral course when the pilot keeps the reticle image centered on a ground target. During this downward spiral course, the dive angle becomes progressively more and more steep. As the dive angle increases, the speed of the aircraft also increases. The increasing dive angle constitutes rotation of the aircraft about its elevation axis. Hence, as the dive progresses, there is an increasing turn rate input that tends to cause the elevation gyro to precess away from the mechanical stop and to overcome the weight of the bombing accelerometer.

At the same time, the increasing speed rate during the dive causes ram air pressure to expand a tubular bellows which also tends to raise the weight of the bombing accelerometer. Somewhere along the spiral path downward, the increasing precessional force of the gyro will overcome the decreasing effective weight of the bombing accelerometer. At this point the gyro rises from the mechanical stop, creating an error signal that causes the differential relay to operate.

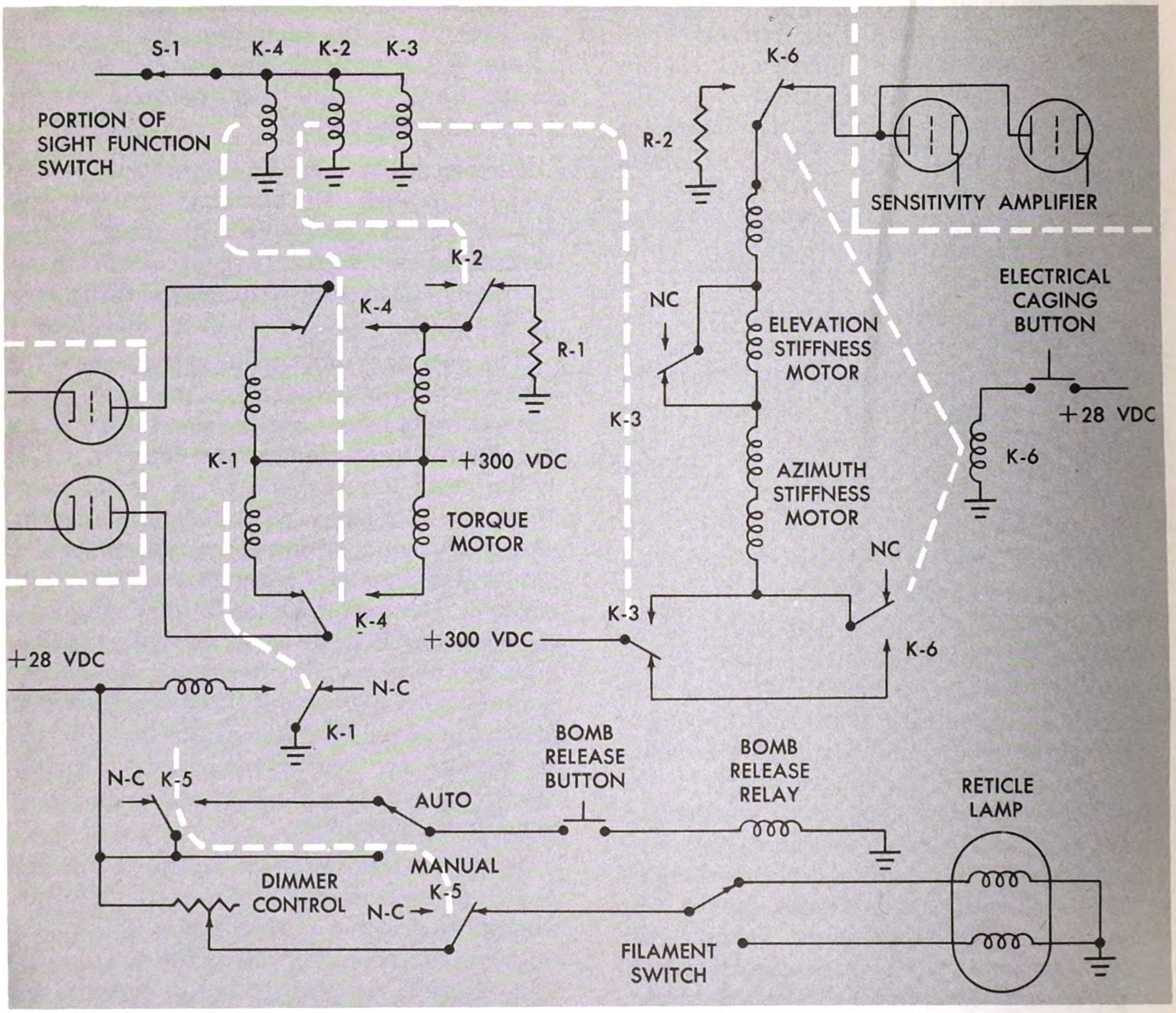
Operation of the differential relay energizes the automatic release relay which energizes in turn the bomb release relay and the bomb rack relays. At the same time, the automatic release relay extinguishes the reticle image to let the pilot know that the bomb release point has been reached.

Electrical Circuits. The electrical circuits which establish the bombing function are shown in the simplified diagram. When thrown to the BOMB position, a portion of the sight function switch energizes the coils of bombing relays *K-2*, *K-3*, and *K-4*.

The energization of these relays causes the contacts of *K-4* to disconnect the windings of the elevation torque motor from the elevation amplifier and to connect the coils of differential relay *K-1* in their place. At the same time, the contacts of *K-2* complete a power circuit through the upper half of the torque motor windings and through resistor *R-1* to ground. The resulting excitation applied in this manner is such as to drive the prediction mirror to its lower mechanical stop. The contacts of *K-3* short out three-fourths of the windings of the elevation stiffness motor and deenergize the stiffness motors by disconnecting the plate voltage of the tubes in the sensitivity amplifier.

When the release point is reached during the bomb run, the elevation gyro precesses enough to create an error signal. The resulting amplifier output energizes the coil of *K-1*, causing contacts of *K-1* to energize the coil of relay *K-5*. One set of *K-5* contacts closes to energize the coil of the bomb release relay, whose contacts (not shown in the diagram) energize the bomb rack relays. At the same time, another set of *K-5* contacts opens to extinguish the reticle image.

During the bombing function, electrical caging is initiated by depressing the electrical caging button which energizes the coil of relay *K-6*. The contacts of *K-6* complete a circuit from the plate supply to ground through stiffness motor windings and resistor *R-2*. The resulting excitation drives the azimuth gyro to its neutral position, but, because only one winding of the elevation stiffness motor is energized, the caging

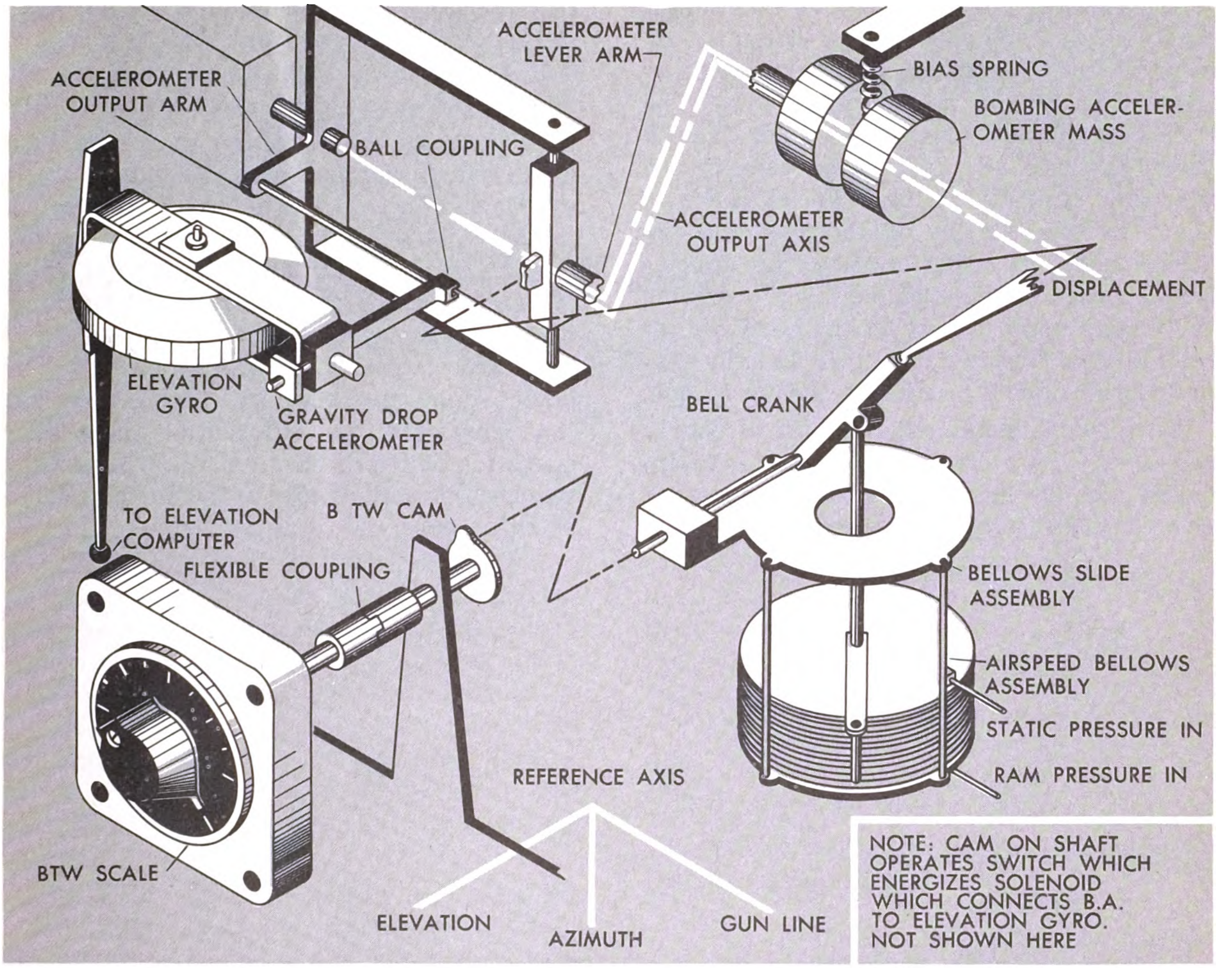


Bombing Electrical Circuits

action is such as to drive the prediction mirror toward its lower mechanical stop.

Bombing Accelerometer. A simplified mechanical schematic diagram of the bombing accelerometer is shown. When the pointer of the BTW scale is turned clockwise away from the GUN-ROCKET position, a cam-operated switch (not shown in the diagram) energizes a solenoid. The resulting operation of this solenoid engages the bombing accelerometer with the elevation gyro through the ball-and-fork coupling. The bombing accelerometer mass may be considered as a weight at the end of a lever pivoted at the accelerometer output axis. The force of gravity, pulling down on this weight, produces torque about the accelerometer output axis. This torque raises the accelerometer output arm. The linkage through the ball-and-fork coupling tilts the elevation gyro against its lower mechanical stop.

The bombing accelerometer is opposed by the tension of a fixed bias spring, which tends to support the mass of the accelerometer. In addition, the amount of torque produced by the accelerometer mass can be



Bombing Accelerometer

varied by changing the displacement of the accelerometer lever arm relative to the accelerometer output axis. Increasing the displacement increases the torque.

One of the factors affecting the displacement of the accelerometer lever arm is the BTW cam. This cam is positioned by turning the pointer to the proper value on the BTW scale. This setting depends on the speed and direction of the surface wind at the target. Positioning of the BTW cam causes the bellows slide assembly to move up or down. The vertical motion of the bellows slide assembly is transmitted through the bell crank to change the displacement of the accelerometer lever arm.

A second factor affecting the displacement of the accelerometer lever arm is the mechanical output of the airspeed bellows assembly. Since the inputs to the airspeed bellows assembly are static air pressure and ram air pressure, the length of the bellows increases as the true airspeed increases. As the airspeed increases, the bellows linkage drives the bell crank so as to decrease the displacement of the accelerometer lever arm. Therefore, as the airspeed increases during the drive, the torque applied to the elevation gyro by the bombing accelerometer will decrease.

As the pilot follows the downward spiral course dictated by the depression of the

reticle image, the curvature of the flight path increases. This curvature causes an increasing rate of rotation about the elevation axis to be felt by the elevation gyro. The direction of this rotation is such that it tends to cause the gyro to precess in the direction opposite to that in which it is tilted by the weight of the bombing accelerometer. In other words, the precessional torque of the gyro begins to overcome the torque applied to it by the bombing accelerometer. However, the gyro actually cannot precess until it produces a greater torque.

During the bomb run, the spiral course causes the gyro precessional torque to increase, and the increasing airspeed causes the accelerometer torque to decrease. When the increasing gyro torque becomes greater than the decreasing accelerometer torque, the gyro will precess away from its mechanical stop, moving the gyro pickoff armature with it.

As long as the elevation gyro was held against its mechanical stop, the output of the gyro pickoff was equal to that of the torque motor pickoff, and the error signal to the amplifier was zero. As soon as the gyro precesses, an error signal is created by the movement of the gyro pickoff armature away from the position of correspondence. The error signal produces an amplifier output that energizes the differential relay. The operation of the differential relay begins the sequence of actions that result in the release of the bombs.

CHAPTER 4

INTERCEPTOR AIRCRAFT AND MISSILE COMPUTERS

At the present time computing devices are a vital factor because of the revolutionary increase in the level of complexity of weapons control systems in contemporary fighter interceptor aircraft. As you know, many problems have been encountered in the development, operation, and maintenance of fighter interceptor aircraft and missiles. Also, tremendous amounts of money and manhours have been and still are expended on their solution.

Research has been greatly aided by the use of mechanical and electronic *simulator* devices for the solution of problems pertaining to flight trajectories of missiles, automatic control, fire control, and guidance systems. By means of these devices which we call computers, many of the most difficult problems are being solved in a fraction of the time that would be required by human effort alone. We may say then that the purpose of computing or information processing equipment is to reduce man's work, or to amplify through automation the work that he has to do. Hence, information processing expands and supplements man's mental efforts in the same manner that the application of power from our natural resources has supplemented his physical labors.

The purpose of this chapter is to acquaint you with the basic principles of computing devices, investigate the elements involved, and point out some of their specific applications in the field of fighter interceptor aircraft and missiles. As you progress into the text of this chapter, you will find computer elements of many types and of widely varying complexity which are incorporated into the circuitry of guidance and control systems. For example, the computer unit in a guidance system may be a simple mixing circuit within the airborne equipment; or it may be very complex, as in a large scale ground installation where a complete flight program may be determined.

DEFINITION AND USES

A computer may be defined as a device which accepts quantitive information, rearranges or translates it into various forms, performs mathematical or logical operations on it, and makes available a form of quantitive information as an output. In short, a computer is a data processing device.

It may be necessary to process and evaluate much complex data in very short intervals of time in order to obtain complete flight information from a single interceptor aircraft or missile launching or from a minimum number of launchings. Certain *telemetered data* must be instantaneously evaluated to insure range safety. Much data must be evaluated so that the findings may be used for correction or change in the operation of an aircraft or missile that is to be

launched later. With such requirements, it is necessary that the data be resolved into some simple form in order that a computer may accept and process information from many channels with a high degree of accuracy. Computer devices have saved much time and expense through their ability to accept, evaluate, record, and store a large amount of data from single launchings of interceptor aircraft and missiles.

In the past, one of the greatest drawbacks involving the use of computers has been in the preparation or *programming* of problem material into a form suitable to the computing device in question. Newer units are now being developed which are able to assimilate information until it is required. These developments have reduced the *set up* time for the computation and, hence, have further lessened the human factor in the computing process.

Although it is possible to design computers which will accept very complex data, such devices are very complex in circuitry. They also require longer periods of set up time than do machines fed input data which has been reduced into simple terms. This has led toward the development of data programming based on simple, well-known principles of mathematics and logic, such as the *binary* and *denary* systems and *Boolian algebra*, which are treated near the end of this chapter.

CLASSIFICATION OF COMPUTERS

Computers may be classified in terms of the phase of an aircraft mission with which they are associated. They may be classified as distinct units or combinations of units, such as:

- Automatic tracking computers
- Angle of attack computers
- Dive angle computers
- Launch and pre-launch computers
- Elevation or azimuth computers

However, these are only a few of the many types in use.

Another method of classifying computers is by their principles of operation. Both airborne and ground-based computers are composed of two types of components, analog and digital. Many of the computers associated with automatic control, fire control, and guidance systems are of the analog type. However, they may be mixtures of both. Analog-digital converters are used to translate data from one form to the other.

Analog Computers

The analog computer manipulates physical quantities which represent mathematical variables of a particular problem to be solved. You are perhaps familiar with such devices as the slide rule, and various types of differential gearing. These devices use computing principles wherein physical quantities are made to follow mathematical laws.

The simplest analog computer is perhaps the slide rule. The slide rule represents logarithms of numbers by measured lengths on its scale, where the distance between graduations representing two numbers is proportional to the difference between the logarithms of the numbers. For example, in performing the operation of multiplication, we select the graduated length corresponding to the logarithm of the multiplicand on one scale and that corresponding to the multiplier on the sliding scale which is moved so that the selected lengths may be added linearly. The sum of the selected lengths is the logarithm of the resulting product, which is read directly from the third scale. By similar processes we may divide, raise numbers to any power, or extract roots of numbers rapidly with a relatively high degree of accuracy.

In a mechanical type analog computer, the machine variables may be rotating shafts driven by gear trains. The angular displacement of each shaft may be measured to produce the solution of an equation or some mathematical operation. In the case

of the automobile speedometer, the indicator is moved either mechanically or by the generation of electrical energy which varies in proportion to the speed of the drive shaft. This operation of indicating the rate of change of the automobile's velocity involves the principles of differential calculus. These principles will be investigated more thoroughly in subsequent sections of this chapter.

The differential gearing in your automobile is a mechanical subtracting device in which the amount of subtraction is determined by the gear ratio of the differential gears. Extra speed gained by one wheel is subtracted from the other. We know when an automobile is turned sharply the outside wheel must travel through a longer arc than the inside wheel, meanwhile the drive shaft is applying the same force to a common ring gear. The differential gearing serves to adjust the driving force in proportion to the distance traveled by each wheel so that the driving force is equalized. Computers for solving navigation, bombing, and radar gun director problems have used these principles for years.

The techniques of analog computations are relatively simple and are well understood. For this reason they are widely used in computers covering a broad range of complexity. The rapid progress in fighter interceptor and guided missile development owes a great deal to the availability of such analog computers as the *Typhoon*. The *Typhoon* computer is perhaps one of the most complex analog computers of its type. It is able to simulate a pilotless aircraft in flight in *real time* to a very high degree of accuracy and completeness.

Real Time and Arbitrary Time Computers

At this time it may be well to point out a special classification of computer problems where the parameters of time are essential in the solution of a problem. When the time represented in a computer is the same as clock time, the computer operates on a 1 to 1 time scale and is said to be operating in *real time*. However, if the time representation is something other than 1 to 1, a choice of the time scale is permissible, and the computer is said to be operating in *arbitrary time*. The arbitrary time scale may be faster than real time, but it is usually slower.

REAL TIME COMPUTERS. Analog equipment may work in what is called real time. That is, it may continuously offer a solution to the problem it is solving, and the solution is appropriate at every instant to all input information which has so far entered the system. It can thus respond promptly to changing input data, and offer an up-to-date solution at every moment. This property of working in real time is very important in most problems of automatic control. For example, an autopilot flying an interceptor aircraft must respond at once to an altitude change resulting from a gust of wind—the most precise information on how to adjust controls will be useless if it comes 30 seconds too late.

ARBITRARY TIME PROBLEMS. Many of the arbitrary time problems are numerical calculations, such as computing the equilibrium temperature distribution in a turbine casing or determining the resonant frequencies of a crankshaft. Although it may be desirable to complete arbitrary calculations soon, the instant the result is obtained is not critical, as is the case in a real time computation.

A time scale factor greater than unity is valuable when a dynamic system consists of rate problems where rates of change are very high. Where rates of change are very high, significant variation of a variable may take place within the lag time interval of some of the elements of the computer. Computer time scales with values less than one are desired in the simulation of systems where the actual dynamic events occupy extremely long time intervals—minutes or hours.

When time scales are changed in DC analog computers, the voltage changes in the

computer remain proportional to the corresponding changes of the mathematical variables. However, the rates at which these voltage changes take place may be speeded up or slowed down in order to improve the accuracy or increase the convenience of a given computation.

You will find that servo driven potentiometers are employed very frequently in electromechanical multipliers associated with interceptor aircraft computers. The design of these devices has been highly developed, and their usage has resulted in almost errorless output. However, the use of electromechanical computing elements limits the speed of operation of some of the computer variables. This imposes limitations on the computer time scale and on the computation speed. Hence, slow time scales are perhaps more desirable where mechanical elements are involved.

Differential Analyzer Simulator

A differential analyzer is an analog device which may be used for simulating a dynamic system involving time, motion, and distance. It is a device utilized to accomplish the solution of rate problems and differential equations. A simulator may be defined as an electromechanical or wholly electronic model of a dynamic system whose elements are so arranged that measurement on the model gives useful information about the system.

As previously stated there are advantages in providing computation rates either slower or faster than unity time rate. The disadvantage of using something other than unity time scale is that the elements of the actual system cannot be inserted or made part of the simulator system. Hence, in some cases you may find the term *simulator* restricted to computers that operate in real time, and the term *differential analyzer* used to indicate simulator computations in arbitrary time. Hence, the description of a particular computer as an analyzer does not necessarily mean that it be restricted to what might be called analyzer methods. Also, equipment that might be called simulators may be used as differential analyzers.

A simulator is so called because it is intentionally constructed as a model of the actual dynamic system under investigation—the variable quantities in the computing apparatus having definite counterparts in the system.

A differential analyzer may be regarded as a black box receiving problems at the input and giving solutions at the output. The elements inside the box are of no concern as long as it produces the correct results. The simulator, however, is an analog computer in which the various computing elements—such as integrators, multipliers, and adders—are chosen and connected together in a special way to represent a particular type of mechanism, such as an automatic pilot control system in an aircraft.

OPERATION AND APPLICATION. Let us consider, in general, how simulators or analyzers may be used in the interceptor field. Analog computers are capable of analyzing flight trajectories and simulating flight characteristics of both the interceptor and the target.

The type of flight behaviour in which we are mainly concerned is perhaps the simulation of an aircraft or missile automatically homing on a moving target with the aid of radar. The object may be to preflight the interceptor by simulation—to study how it behaves when chasing its target, which may be taking violent evasive action, and to see if it does in fact make a successful interception.

In order to simulate these conditions, the analyzer must have many computing sections which will represent the behavior of the target and all the various processes of the interceptor control operations including the radar system.

Each computing section handles a particular phase of the complete problem. The major sections are composed of the flight simulator,

aerodynamic computer, target and guidance computer, and recording and display devices. In the use of displays the simulator can be made to assume and display attitude and control surface displacement. A trajectory display can simulate and show the instantaneous attitude of both interceptor and target in three dimensions.

The target computer generates a trajectory and simulates target positions in the form of earth coordinates. Target maneuvers are further simulated by changing values of velocity, climb, and turn. Plotting units may be used for plotting interceptor and target trajectories in both horizontal and vertical planes. In fact, a multitude of desired variables may be recorded by certain types of photoelectric records.

A flight simulator and target simulator supply information to the guidance computer, the output of which is fed to an aerodynamic computer. Next, the aerodynamic computer receives information directly from the flight simulator, representing altitude, velocity, and attitude. Then the guidance computer introduces the proper *roll, pitch,* and *yaw* displacement information.

As already mentioned, the resulting maneuvers of both interceptor and target may be traced automatically by moving pens on corresponding plotting tables. Obtaining the necessary signals to move pens and actuate photoelectric recorders involves a long sequence of digital as well as analog operations. Voltage pulses representing output from the interceptor's automatic pilot may be fed to computing elements which correspond to servos driving aerodynamic control surfaces. In turn, signals equivalent to control surface deflections themselves may be derived from which angular and linear accelerations of the interceptor's pitch, roll, and yaw axes are computed. Integrating elements (see Integrating Components of Computers in the third section of this chapter) may then translate these voltages (representing acceleration values) into voltages representing velocities. These velocity voltages may then be converted into signals representing speeds relative to ground or some other predetermined point—perhaps in space. Subsequent integrators may next translate these relative velocity signals into actual airframe displacement signals in roll, pitch and yaw. These resulting interceptor signals may, in turn, be used to control direct recording equipment such as pens on plotting tables—perhaps of the Esterline-Angus Operation Recorder type. (See Output Units in Section F of this chapter.)

In a similar manner, the above processes may be simultaneously accomplished to record target motion and position.

If an interceptor is assumed to be controlled from ground based radar, control signals—translated from the three-dimensional displacement information given by the direct recording equipment (for both interceptor and target)—may be fed directly back to the simulated automatic pilot, because all information is relative to ground. However, if airborne radar is assumed, all information may be relative to the interceptor itself or some other point in space, and the command control signals must be computed by other means.

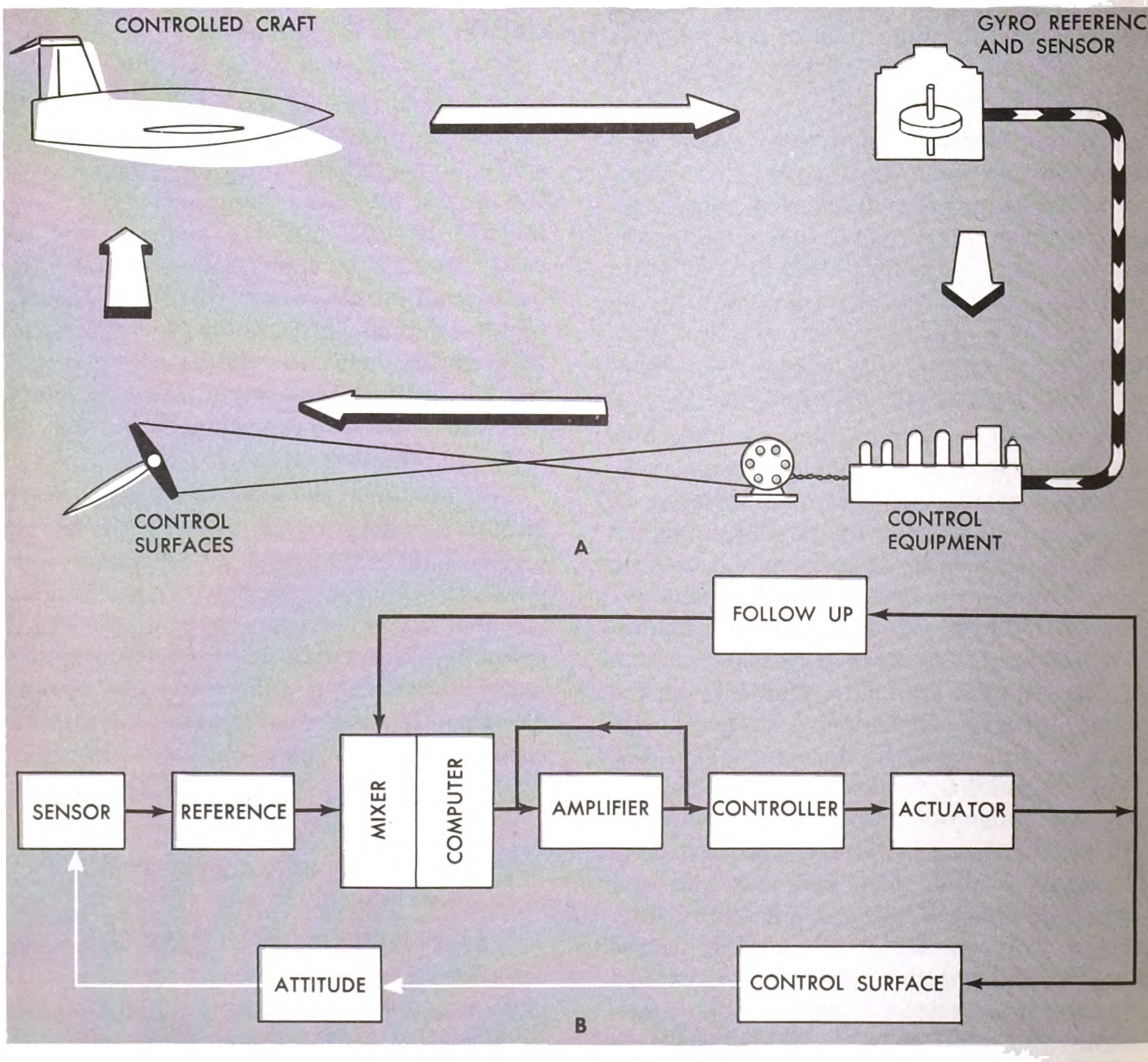
You can visualize the complexity of the circuits involved by considering the inputs and outputs of such a flight simulator computer. In subsequent sections of this chapter, you will find how inputs and outputs in the form of mechanical, hydraulic, and pneumatic linear and angular forces as well as electrical signals may represent variables such as:

1. Linear velocities along a line of flight axis.
2. Angular velocities and displacement about roll, pitch, and yaw axes.
3. Aerodynamic forces along roll, pitch, and yaw.
4. Relative interceptor position with respect to earth or a point in space.

SECTION A, MIXER COMPONENTS OF COMPUTERS

A very common function within a computer unit is the mixing of signals from the sensor and reference units in order to produce error voltages. A sensor unit may be a device which is capable of detecting deviation of an interceptor aircraft or missile from a desired flight attitude. The sensor and reference units may include gyros, airspeed limiters, transducers, and accelerometers.

In order to do its part in, say, an attitude control system, a sensor unit must be able to set up fixed reference lines in space from which a variation or change in attitude may be measured. The sensor must also be able to provide a means of measuring the magnitude and direction of the missile or aircraft deviation from the established reference lines. Finally, it must provide a way of translating the amount and direction of



deviation as well as the amount and direction of control surface movement into a single operation of a complete system. The four basic units of a complete attitude control system are illustrated at **A**. The accompanying block diagram at **B** shows the functional relationship between the mixer, computer, and associated units of a complete attitude control system.

The output of a sensor may be a signal that represents an existing deviation from a desired position. Other conditions such as altitude, airspeed, angular acceleration, or forward and sideward accelerations may also be detected. Although the output of a sensor represents a condition which must be corrected, it does not go directly to the actuator which positions the control surfaces. Notice that the error signal is first fed to the mixer stage of the computer.

Next, the error signal is sent to an amplifier where it is boosted to a value high enough to operate the controller. A controller is a device which uses a signal to vary the energy supplied to the actuator. In the case of an all-electronic system, the controller will be a power amplifier. However, the type of controller utilized will depend upon the type of actuator it must control. As you can see, the actuator will be directly influenced by the controller. The actuator may be electric, hydraulic, pneumatic, or a combination of these. Controllers and actuators will be investigated in the next chapter.

REQUIREMENTS OF MIXERS

A mixer is any circuit or device designed to combine information from two or more sources. Every control system which operates as shown in the functional block diagrams (previously mentioned) must combine several signals. Reference to diagram B shows that a follow-up signal is combined with the sensor signal. An integral or rate signal, which may be produced within the computer, may also be combined with the sensor and reference signals. This mixing occurs when the required aerodynamic reaction depends on more than one variable. Suppose, for example, a system is to control the pitch of an aircraft or missile. This pitch could be influenced by a change in airspeed, attitude, or altitude. Hence a separate sensor and reference unit is necessary to detect each of these three variables. Then all of these signals must be mixed so that each variable will have the required influence on the pitch channel.

Scale Factors

In order for a mixer to function properly it must combine the signals in the correct amplitude proportion and sense of direction. That is, the sense of the follow-up signal must be such as to have a counteracting effect on the output. The total output, in turn, must have the proper amplitude for the correct amount of control. Normally the signals either add or subtract, although other functions may be performed.

If a feedback signal is to be combined with an attitude error signal, the feedback must be of a certain strength with respect to the attitude error signal. This correct proportion may be accomplished by utilizing electronic or mechanical components with the proper weighting or scale factors which determine the comparative influence of a signal. For example, suppose portions of the three signals e_1, e_2, and e_3 are to be added. If only one-third of e_1 and one-half of e_2 are to be used as compared to the strength of e_3, one-third and one-half would be the scale factors of e_1 and e_2 respectively.

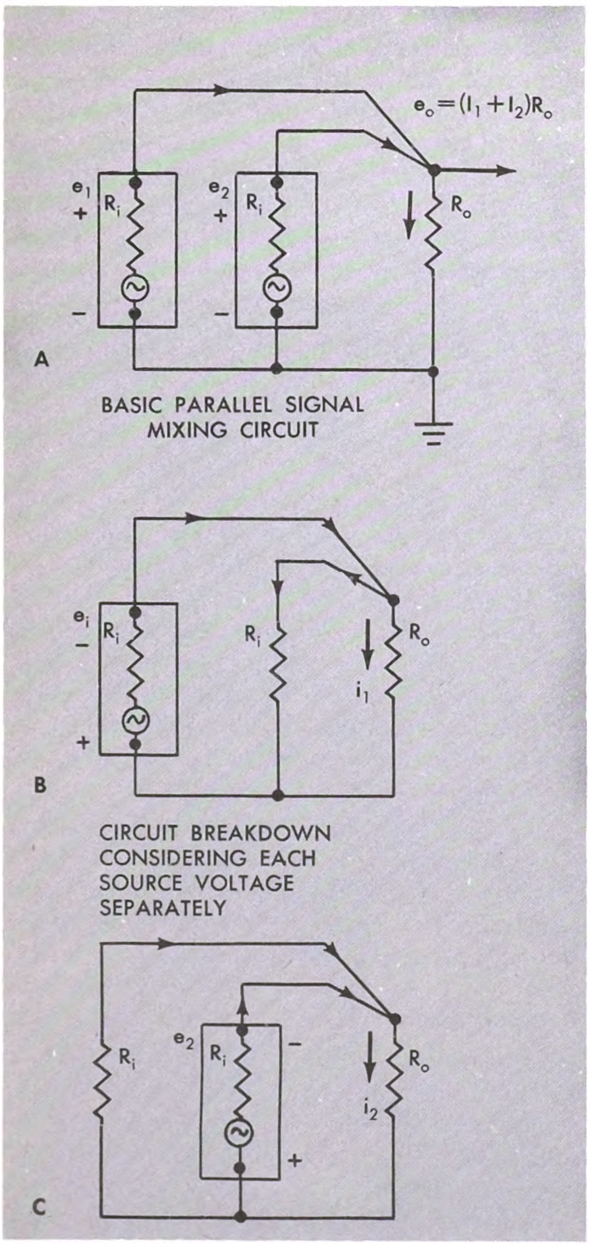
The type of mixer used will depend on the type of control system. Since most control systems in the interceptor field are basically electronic, the mixers discussed in this section will be of the electronic type.

Electronic Mixers

Fundamentally, the operation of all electronic mixers may be reduced to basic network and vacuum tube theory since they consist of either an impedance network of resistors, inductors, and capacitors, and/or several vacuum tube stages. The information

to be combined is represented by the phase and amplitude of several voltages. The voltages representing variables may be from pickoffs, gyros, integrators, or sensors as previously mentioned.

Electrical signals in a control system are usually AC. These signals are then changed to DC before they are applied to the actuators which drive the control surfaces. The phase of the AC signals determines the sense



BASIC PARALLEL SIGNAL MIXING CIRCUIT

CIRCUIT BREAKDOWN CONSIDERING EACH SOURCE VOLTAGE SEPARATELY

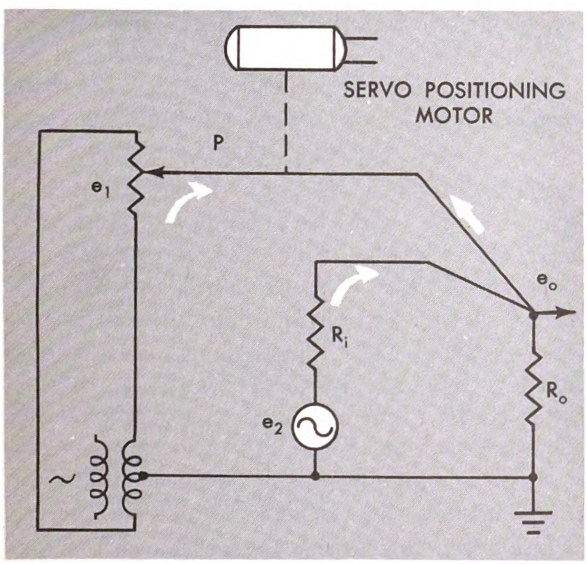
or direction of movement applied to the control surfaces. If two signals are to produce the same effect on a control surface, they must be in-phase in the mixer. It is often necessary to use capacitors and inductors ahead of the mixer in order to alter the phase of incoming signals so they will mix in exactly the correct phase. If DC signals are used they must be mixed with proper regard to polarity.

PARALLEL MIXER. The simplest way to mix two source voltages is to combine them in parallel across a common impedance R_o as shown at **A** in the next illustration.

It is more difficult to visualize the operation of a circuit of this type since it has more than one source. You may recall, however, that a good method of analyzing a circuit having more than one source voltage is by the use of the superposition theorem. For example, if a network of linear impedances is energized by two or more sources, the current through or the voltage across any specific resistance (R_o in our example) can be expressed as: the sum of the currents or voltages that each source would produce—were it alone connected in the network—each remaining source being replaced by its internal impedance.

Investigating the accompanying breakdown of the basic parallel mixer may help you in visualizing the currents and voltages produced by each source separately. The current flow indicated by the arrows in the illustration results from assuming an instant when the polarities of the AC sources are as shown. In each case of the breakdown (at **B** and **C**) the currents i_1 and i_2 represent the output currents due to a single source voltage. When the original circuit at **A** is considered with both voltage sources, the actual current through R_o is the vector sum of i_1 and i_2—the components of current resulting from the sources considered separately. Since the total current is the vector sum of the branch currents, the output voltage e_o will depend on the internal impedance of each source as well as the strength or magnitude of its voltage.

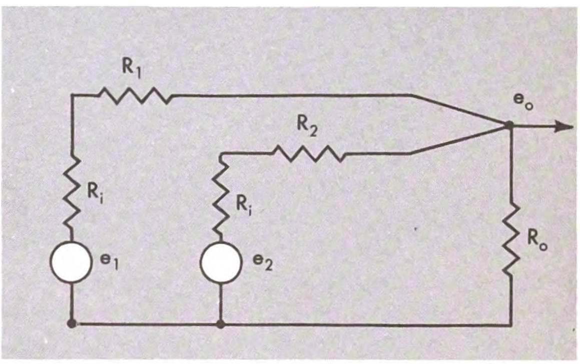
Changes in internal impedance of pickoffs which are interconnected may result in signals of incorrect amplitude. For example, let us consider the circuit where the generator e_1 of the basic parallel mixer is replaced with a potentiometer pickoff as shown. The current supplied by generator e_2 will depend on the total impedance in the circuit. When this impedance is altered by movement of the potentiometer pickoff arm, the current will vary, thus causing the voltage drop across the internal resistance of e_2 to change. Consequently, the voltage applied to R_o from e_2 will be altered even though the change does not represent a change in the information represented by e_2. Of course, we expect some change in voltage across R_o due to the change in current supplied by e_1. It is desirable that this change be proportional to the change in the potentiometer pickoff.



Parallel Mixer Utilizing a Pickoff Potentiometer with a Variable Internal Impedance

In order to limit the undesirable change in the current supplied by e_2 when the pickoff for e_1 is varied, a limiting or attenuator resistor may be inserted at point P between the pickoff arm and R_o.

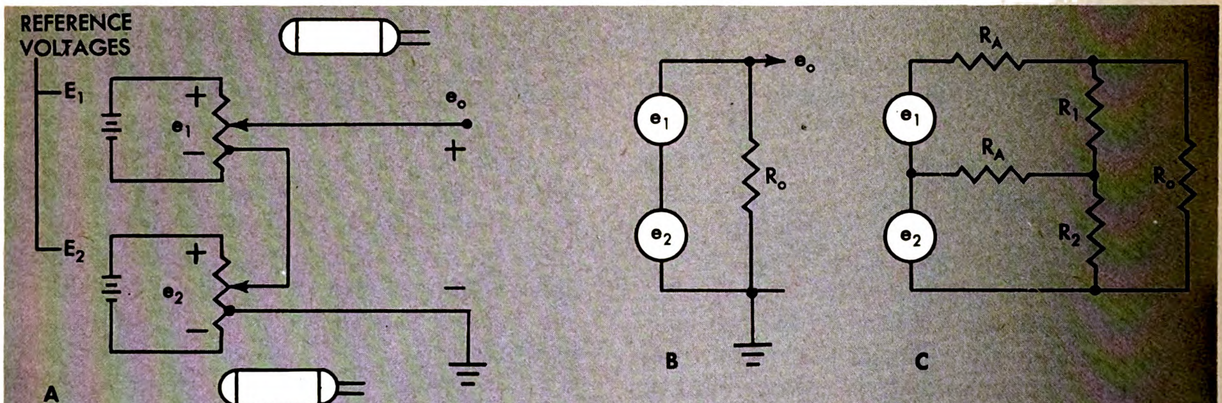
Attenuating Resistors are normally inserted in series with each individual source as shown in the next illustration. These series resistors will reduce errors caused by coupling between the signal sources. This coupling is produced by current from one source changing the voltages which would normally be produced by the other source. Such an error occurs from a source whose internal impedance varies, such as in a potentiometer pickoff when its pickoff position is altered.



Parallel Mixing Circuit with Series Scale-Factor Attenuating Resistors

Secondly, the resistors allow the designer to preadjust the amount of the signal which should be applied to R_o. This will enable the signals to be mixed in the proper ratio (scale factor), which is very important. Often, it is easier to alter the magnitude of a pickoff output by using an attenuating resistor of the proper value than it is to acquire a pickoff with the proper output. The use of attenuating or scale-factor resistors in electronic summing networks will be investigated in subsequent sections of this chapter.

SERIES MIXERS. Variables appearing as shaft rotations may be mixed (summed up) with the use of simple series electrical circuits by first converting the shaft rotations into equivalent voltages or currents. Notice, in the example illustrated at **A**, the voltages e_1 and e_2 (representing the variable to be added) are set to their particular values by the mechanical rotation of potentiometer



SERIES MIXING CIRCUIT USING SERVOS FOR POSITION PICKOFFS

EQUIVALENT CIRCUIT WITH AND WITHOUT ATTENUATING RESISTORS FOR DECOUPLING

shafts. Upon further examination of the circuit you will find that the center tap in each case establishes a reference for the polarity of e_1 and e_2. Hence, these voltages may be either series aiding or subtracting, depending on which side of the center tap the pickoff potentiometer is set. Assuming an open circuit between the output terminal and ground, the voltages e_1 and e_2 will be directly proportional to the setting of the pickoff brushes. From Kirchhoff's law, it is apparent that the output voltage is equal to the algebraic sum of e_1 and e_2.

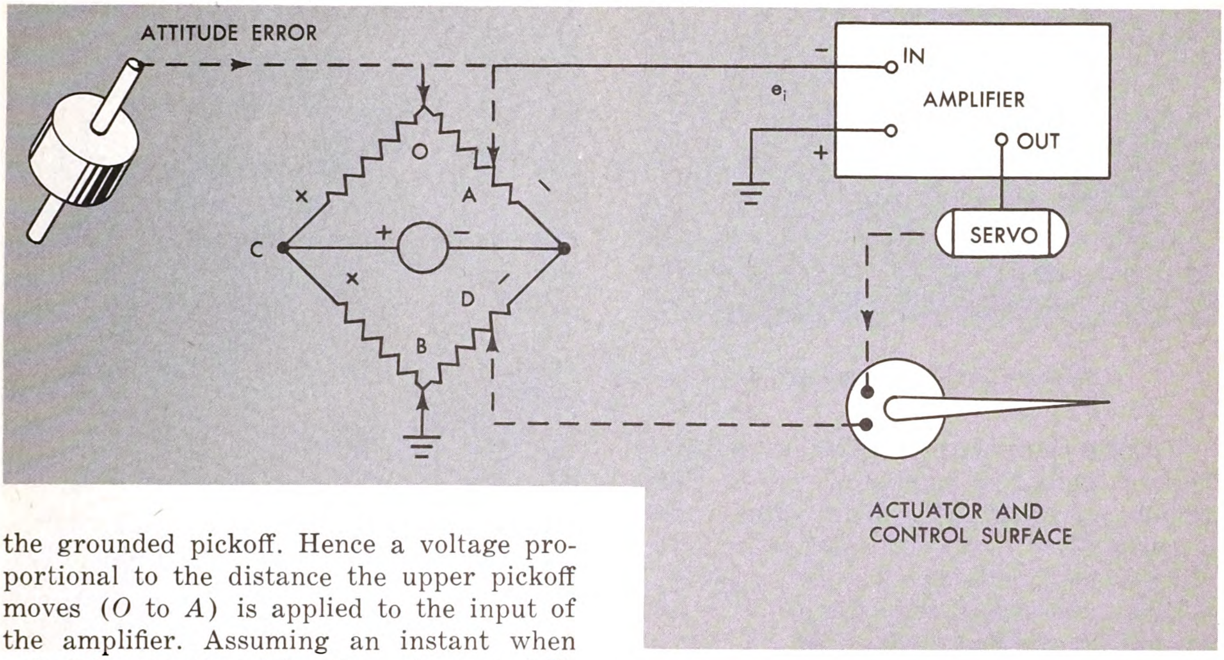
Rather than providing direct current sources for the reference voltages E_1 and E_2 as at **A** in the above discussed illustration, AC sources of identical phase may be used. The voltages on either side of the center tap will still be in phase opposition, allowing the algebraic summation of the instantaneous values of AC.

As shown in the equivalent circuit at **B**, the voltages e_1 and e_2 are combined by applying them across a common load R_o. Since all the load current must flow through each source, this system is normally altered as shown in the equivalent circuit at **C**. Here the voltage across R_o is equal to the sum of the voltages across R_1 and R_2. The attenuating resistors, designated as R_a, are included for decoupling as in the case of the parallel mixers previously discussed.

RESISTANCE BRIDGE MIXER. Position information from two sources may be combined with the use of a resistance bridge network as shown in the related illustration. In this example, the position information from a gyro is detected and combined with the position information from the control surface. That is, a voltage proportional to gyro displacement is mixed with a voltage proportional to the control surface position.

The bridge as illustrated will serve the purpose of two electrical pickoffs and a mixer. Note that two sliding contacts are used in the bridge; one contact is mechanically connected (as indicated by dashed lines) to the gyro gimbal and the other to the control surface. Thus, physically, the potentiometers may be located quite some distance apart. The potentiometer pickoffs are adjusted so that when the interceptor (aircraft or missile) is in the proper attitude and the control surface is streamlined, the bridge is balanced. Therefore, there is no difference in potential between points *O* and *B*; consequently, no signal will be fed to the amplifier to effect a correction.

If the attitude of the craft deviates, the upper pickoff contact will move along its potentiometer (say to the right toward *A*). Thus, the bridge will be unbalanced, producing a difference in potential between *A*, the new position of the upper pickoff, and *B*,



Resistance Bridge Mixing Circuit

the grounded pickoff. Hence a voltage proportional to the distance the upper pickoff moves (O to A) is applied to the input of the amplifier. Assuming an instant when polarities (produced by the reference AC) are as shown, the gyro displacement component of input voltage to the amplifier will be negative with respect to ground.

The output of the amplifier is fed to the *control servo* and *actuator* which positions the control surface to bring about the required attitude correction. At the same time, the direction of the actuator rotation is such that the lower pickoff wiper is moved to the right, producing a positive follow-up signal proportional to the distance the lower pickoff moves from B to D—which is also applied to the input of the amplifier. Actually both actions occur at the same time, and the voltage applied to the amplifier will be the resultant of the original negative error voltage and the positive follow-up component.

In this example, the bridge mixer performs a subtraction since the signals caused by the displacements tend to cancel out. This can be shown by the voltage equation for the closed loop. Starting at the grounded input terminal

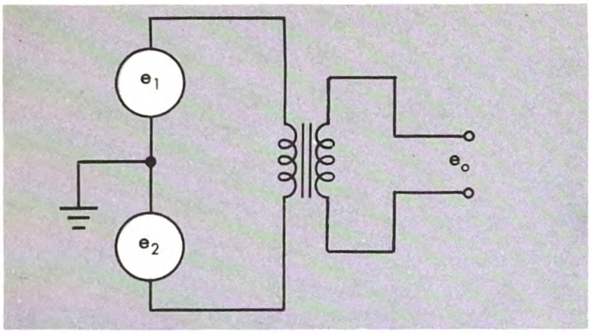
$$e_i = -AC + CD$$

The resultant input e_i will be negative since the deflection of the upper pickoff arm is greater.

The scale factors will depend on the mechanical linkages and on any decoupling or attenuating resistors which may be added to the bridge circuit.

For a more detailed step-by-step discussion of the operation of a bridge mixing circuit and its interaction between a vertical flight gyro, amplifier, and actuator (resulting from an attitude deviation), refer to *pitch control* in the index.

TRANSFORMER MIXER. The output resistor R_o in the series and parallel mixers previously discussed may be replaced with the primary of a transformer as shown in the accompanying illustration. This is a basic series-mixing circuit with the voltages representing variables connected in series-opposition. The signals will oppose each other in the primary of the transformer since their individual current components flow in opposite directions. Or, if you prefer, the lesser voltage will "buck out" part of the greater voltage. Consequently, the secondary output voltage will be proportional to the algebraic vector sum of the input voltages.

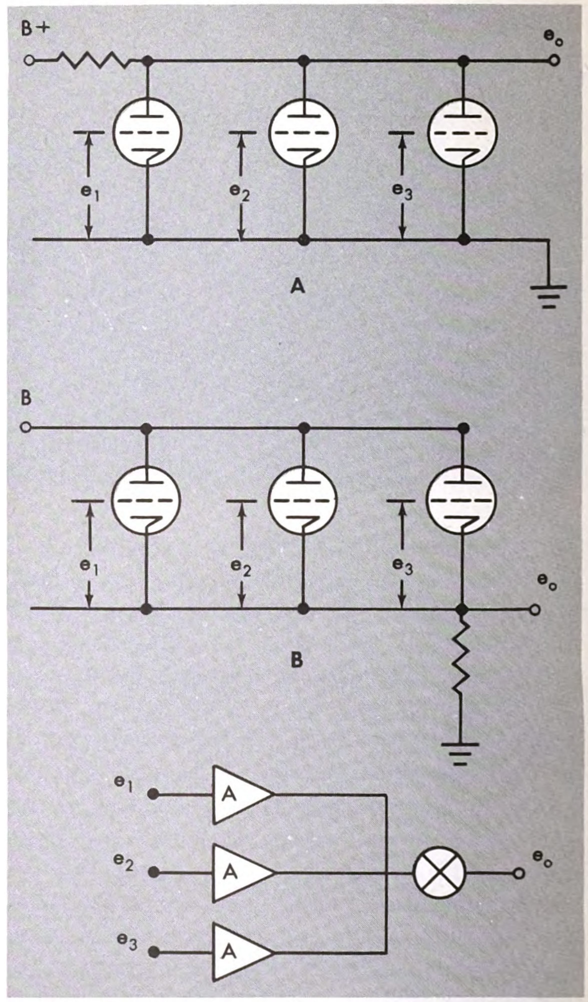


Series Opposing Transformer Mixer

VACUUM TUBE MIXERS. Another very important method of combining one or more voltage sources is by means of vacuum tube circuits. Two typical vacuum tube mixer circuits, sometimes referred to as *parallel vacuum tube adders,* and their symbolic representation are illustrated.

The parallel mixer circuit shown at **A** is of the conventional amplifier type with a common plate load impedance, where the output voltage e_o is proportional to the sum of the input voltages. Of course, the amplifiers also perform the useful purpose of increasing the signal strength. The parallel mixer shown at **B** is of the cathode follower type with a common cathode resistance. Cathode follower type parallel mixers have the disadvantage of not being able to amplify; however, they have the advantage of having very high input impedance and low output impedance. This makes them very desirable for matching the high output impedance of a preceding circuit to a low impedance. When a cathode follower is connected to any type pickoff or sensor, it presents a comparatively light load.

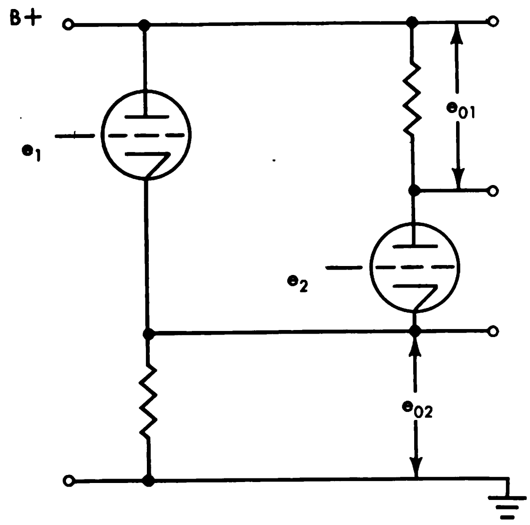
The number of inputs that can be fed to vacuum tube mixers is not necessarily limited. Assuming the tubes within the mixer are identical, similar inputs will have the same influence on the output. Scale factors may be introduced by the use of attenuating resistors or potentiometers at the inputs or in the separate plate or cathode circuits. Also, see *Vacuum Tube Adders.*



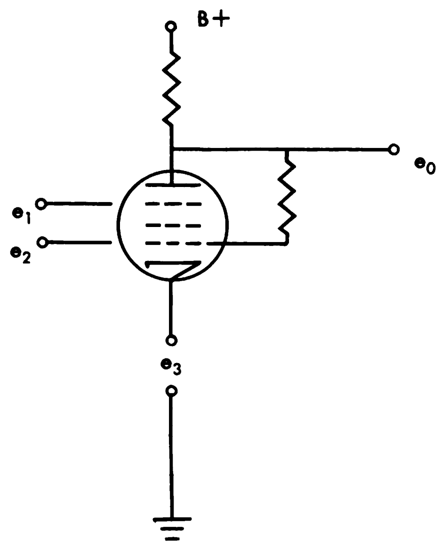
Parallel Vacuum Tube Mixers

The Differential Amplifier Mixer circuit shown in the next illustration is a variation of the parallel amplifiers so far discussed. In this circuit, the amplifiers are altered to produce either a sum or a difference output. That is, e_{o2} is proportional to the sum of the two input voltages, and e_{o1} (the voltage between the plates) is proportional to their difference.

Multigrid Mixing is another effective method of vacuum tube mixing, where several signals may be combined within the electron stream of the vacuum tube itself. Several signal voltages may be impressed



Differential Mixer



Multigrid Mixer

upon the various electrodes of a tube in order to modulate its plate current. The effect of plate current variation due to a signal has little effect on the source of that signal since the impedances between the various electrodes of multigrid tubes are very high. The number of inputs is usually limited by the number of electrodes as is illustrated in the basic multigrid mixer circuit. The scale factors depend on the characteristics of the tubes used and the elements to which the input voltages are applied. Precautions must be taken in order to avoid errors due to the nonlinear characteristics of the tube.

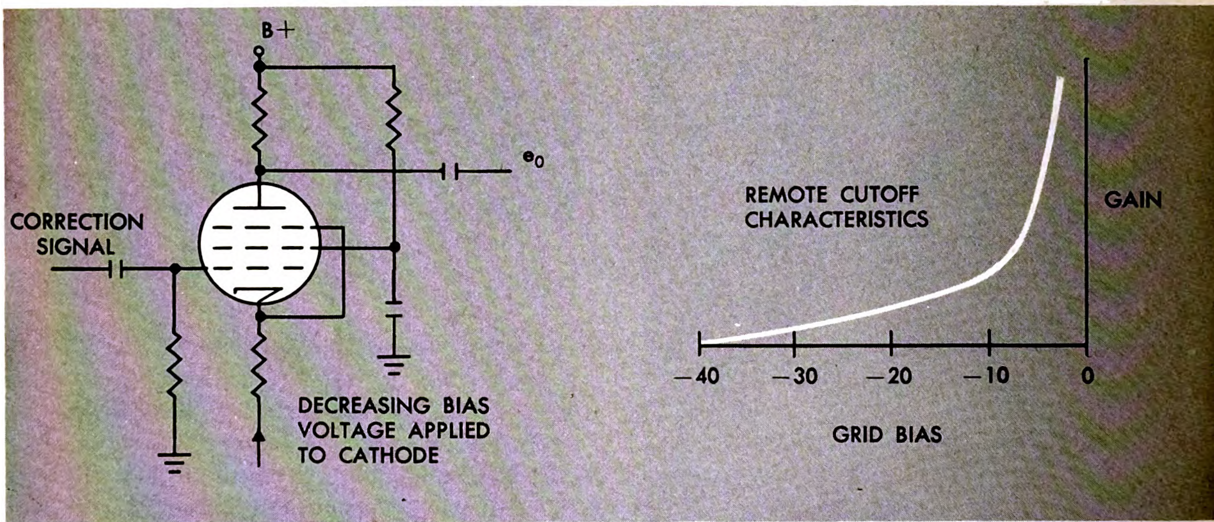
Often such mixing circuits are utilized for mixing signals of different frequencies in communication type electronic equipment such as radios, television, and interceptor missile guidance transmitters. Such mixing circuits are also utilized in mixing the low frequencies used in various attitude control systems.

Mixing Amplifiers are also used whenever it is desired to impress AC or RF pulses upon a varying DC signal as in many telemetering systems. They are used to control amplification of correction voltages transmitted to guidance and control systems when it is desired to increase the sensitivity of a system, such as when an intercepter aircraft or missile approaches its target.

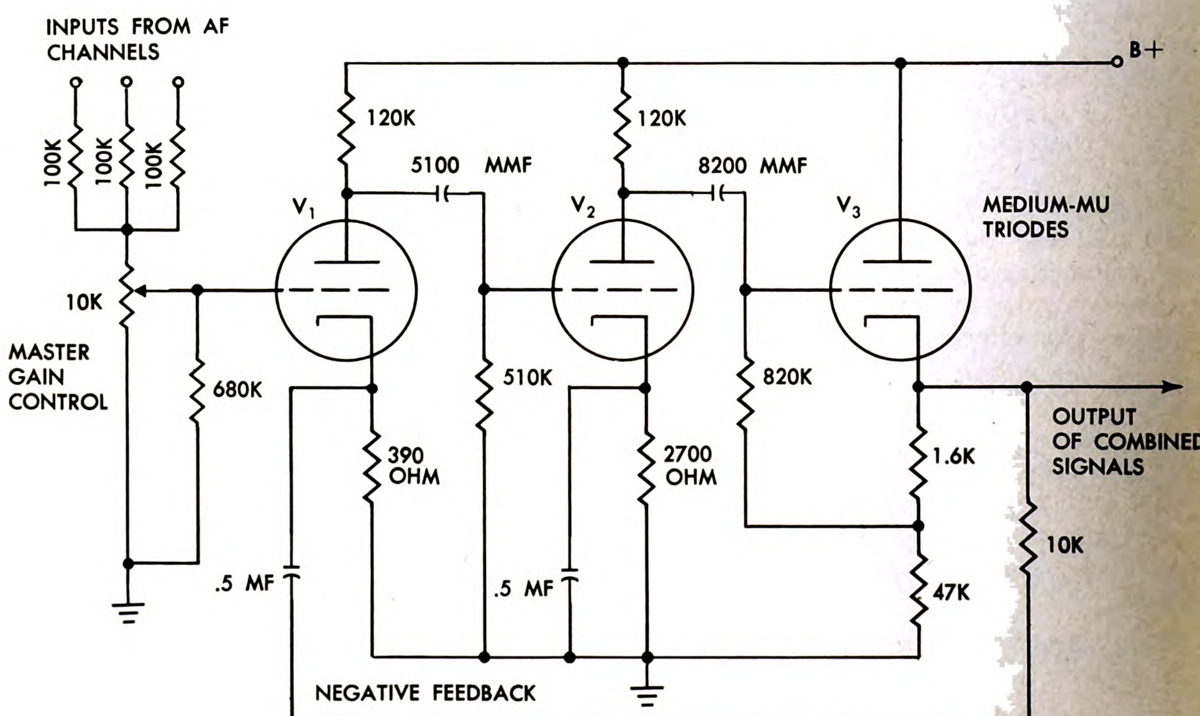
The next circuit diagram illustrates a mixing amplifier circuit in which a decreasing positive voltage is applied to the cathode of a variable mu pentode. You may recall that a decreasing positive voltage applied to a cathode produces the same effect as applying a decreasing negative bias at the grid. Thus, the plate voltage in the circuit will increase as the bias is decreased, resulting in greater amplification and greater sensitivity in the circuit for correction signals.

Another mixing amplifier, as used in some telemetering systems to combine several frequency-modulated signals, is illustrated. The 10K potentiometer serves as a master gain control for the entire circuit. The signals are fed into the amplifier stage by means of a summing network composed of three 100K resistors as shown. The mixed signals are amplified by two stages of conventional RC-coupled amplifiers. The cathode follower output stage (V_3) provides a low impedance output. Negative feedback from the cathode follower output is applied to the cathode of

V_1 through a series RC network in order to reduce the harmonic distortion to a low level, thus providing a flat frequency response for the desired range of frequencies.



Mixing Amplifier



Mixing Amplifier Used in a Telemetering System

SECTION B, MATHEMATICAL COMPONENTS OF ANALOG COMPUTERS

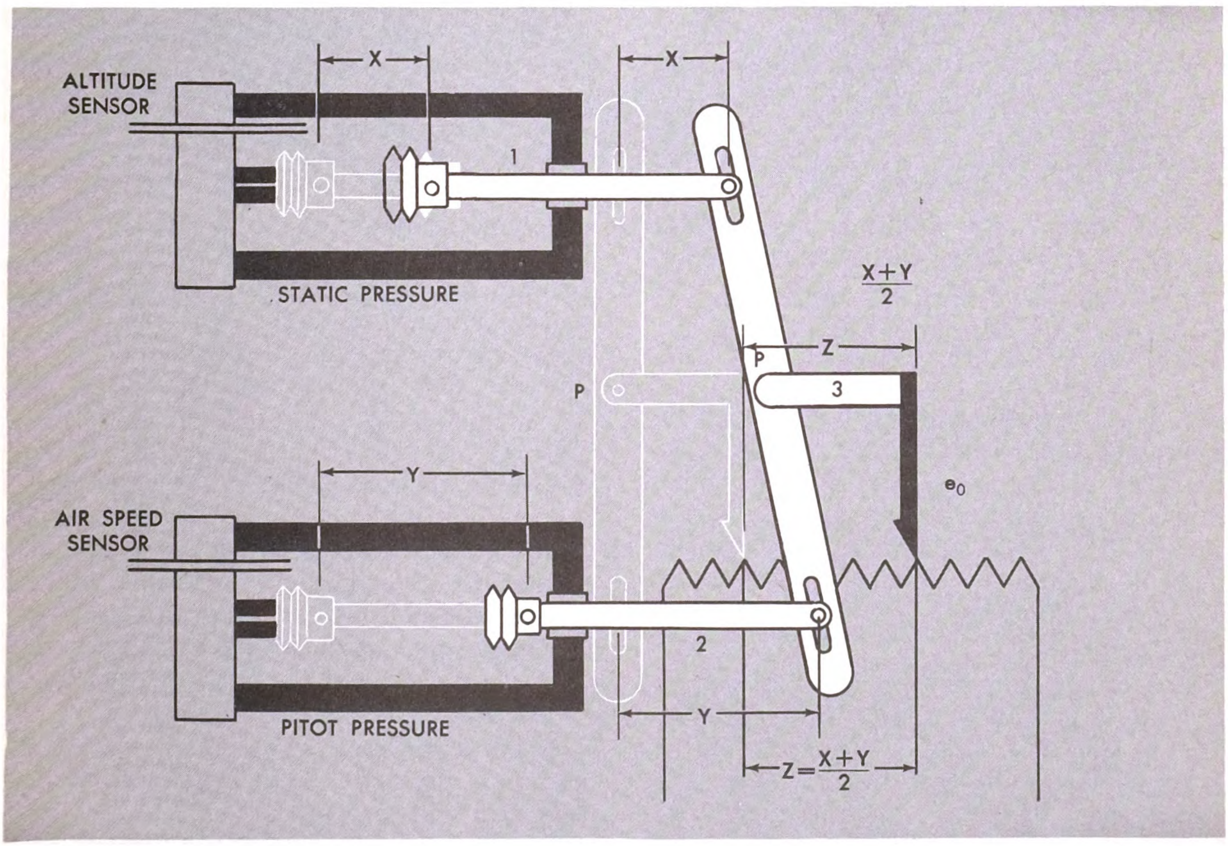
The analog machine is just what its name implies, a physical analogy to the type of problem its designer wishes to solve. Information is supplied to the machine in terms of the value of some physical quantity, such as an electrical voltage or current, the degree of angular rotation of a shaft, or perhaps the compression of a spring. The machine translates this physical quantity in accordance with the rules of its construction. Since these rules have been chosen to simulate the rules governing the problem, the resulting physical quantity is the answer desired. If the analog equipment is used as a control device, the final physical quantity is applied to exercise this desired control.

MECHANICAL ADDERS

The variables to be added and their sum may appear as translations of specified points in a mechanical adder. They may appear as angular displacements or turns of a shaft or as a lateral displacement of shafts and levers.

Linkage Differential Mixer and Adder

One common form of a mechanical adder is the linkage and differential mixer and adder illustrated. Note that shaft *1* is connected to an aneroid diaphragm vented directly to static pressure and shaft *2* is



Linkage Differential Mixer and Adder

connected to a similar diaphragm which is vented to pitot pressure.

Let us assume that shafts *1* and *2* respond independently *in real time* to instantaneous changes in static and pitot pressures and that the instantaneous positions of the shafts represent the information to be added.

The illustration indicates that shaft *1* is displaced x units laterally due to an instantaneous increase in static pressure and shaft *2* is displaced y units because of an increment increase in pitot pressure. The combined displacement of shafts *1* and *2* is transmitted by mechanical linkage and lever to a third shaft. Notice that the lever is connected in such a way that all the mechanical linkage involved may pivot freely and reposition shaft *3* so that its displacement z may represent the weighted summation of the two sensing shafts.

$$z = \frac{x + y}{2}$$

You will note that the direct sum of the two variables represented by $x + y$ is not obtained directly. Upon further examination of the illustration you will find that if the displacement of shaft *2* is made equal to that of shaft *1*, the resulting displacement of shaft *3*, will also be equal to that of shaft

1. That is, $x = y = z$, or $\frac{x + y}{2} = z$; and z

actually equals the weighted average of the individual displacements. This is due to the geometry of the mechanical linkage, which effects an output weighting or scale factor of 1/2. The scale factor in this example could be changed by altering the lengths of the vertical lever arm on either side of the pivot point P.

The output from shaft *3* could reposition an electric pickoff, as shown in the illustration, so that the output voltage e_o would be proportional to airspeed and altitude. The mixer lever eliminates the need for two pickoffs to convert the shaft positions into an electrical signal before mixing.

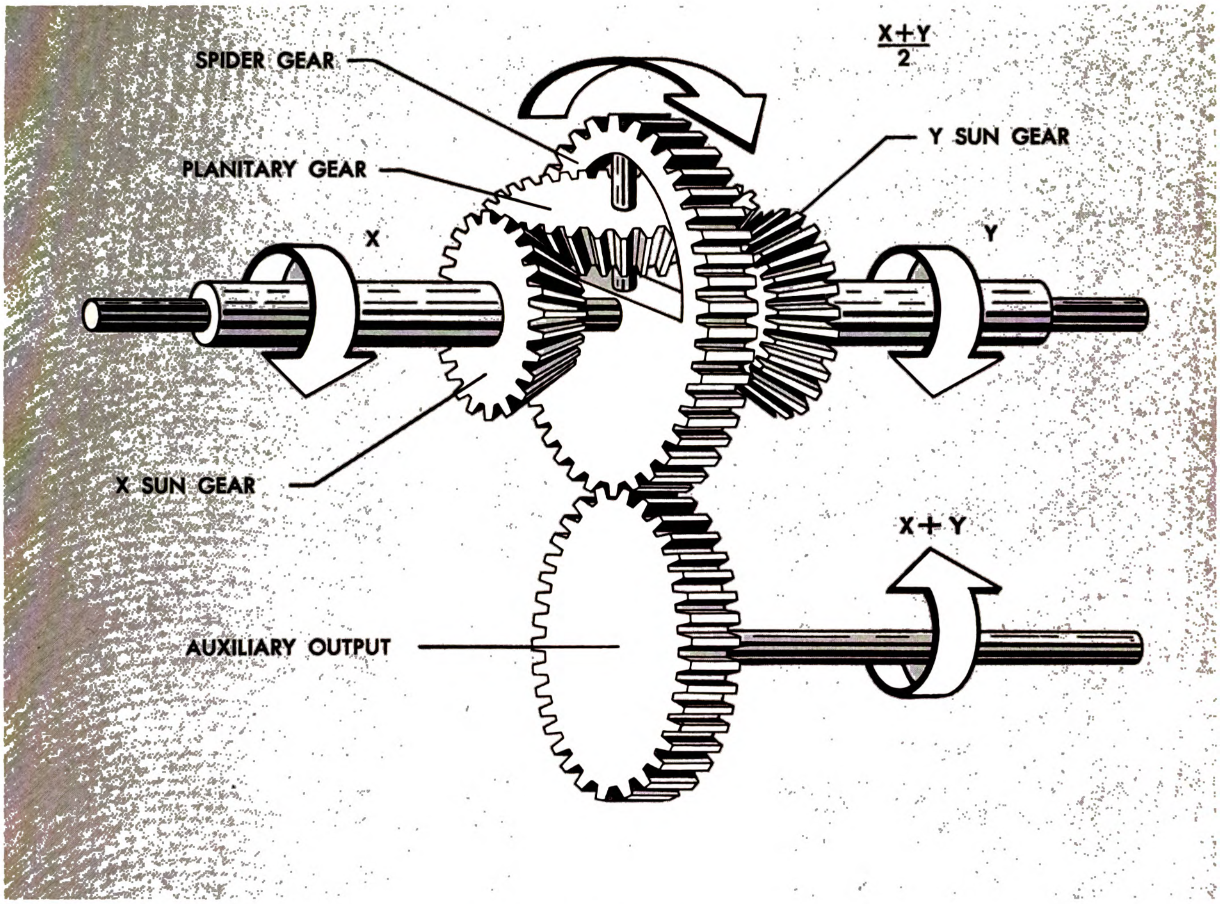
If the inputs to shafts *1* and *2* provide enough pressure, the output could operate a pneumatic or hydraulic valve directly. This would not be a very practical application, since extreme pressures would be necessary, requiring a much more rugged structure. Actually, the best method of construction for a device of this type would be to have the entire mechanism inclosed in a sealed case.

Gear Differential Adders

Gears may be used to combine position or angular velocity information by means of the standard gear differential similar to the differential housing of your automobile. If the input shafts contain position information, the position of the output shaft will constantly indicate the difference between the two shaft positions, as pointed out at the beginning of this chapter. If the information is represented by the speed of the rotating shafts, the angular velocity of the output shaft will represent the difference between the two input velocities. The input shafts may be selected so that the output represents the sum of the inputs rather than their difference. This applies to both position and angular velocity information. Whether a differential adds or subtracts depends upon the sense of the inputs; if both input gears turn in the same direction for increasing values of input, the differential adds. There are many variations of differential gears; however, only one need be considered.

The operating principles of mechanical addition are shown in the accompanying illustration of the *Spider Differential Gear Adder*. The shaft to which the spider gear is keyed is shown to be extending to the extreme left and right—through the collars (shafts X and Y) of the two sun gears. The sun gears mesh with the *planetary gear* and the *spider gear* meshes with the *auxiliary output gear*.

If the Y sun gear is held stationary and X is turned clockwise, as indicated, the planetary gear will revolve on the Y sun gear and cause the spider gear to rotate through one half the number of revolutions that X is



Spider Differential Gear Adder

rotated. In the same manner, turning Y through one revolution while X is held fast, causes the spider gear to turn one-half revolution. Therefore, when both sun gears are turned in the same direction, the spider gear is caused to rotate by an amount equal to one-half their sum.

Now, since the diameter of the spider gear (in the example illustrated) is assumed to be twice that of the auxiliary gear, the output shaft will turn two revolutions for each revolution of the spider. In this example, we see that one input shaft is turned by an amount x, and another input shaft is turned in the same direction by an amount y. Consequently, the output shaft turns by an amount equal to $x + y$.

Turning the input shafts in opposite directions will produce subtraction. In such a case, the output shaft will turn by an amount equal to the difference in inputs, in the direction of the input shaft with the greater amount of turning.

ELECTROMECHANICAL MULTIPLIERS

Although many mechanical computing elements have the advantages of dependability and ruggedness, they have a lack of flexibility in making changes and difficulty in assembly because of excess bulk and weight. Electromechanical computing elements are often more desirable because of their light weight and compactness which leads to ease of assembly. They also have a

short reaction or response time (rapidity of operation) and are often more adaptable to extreme changes in problem characteristics.

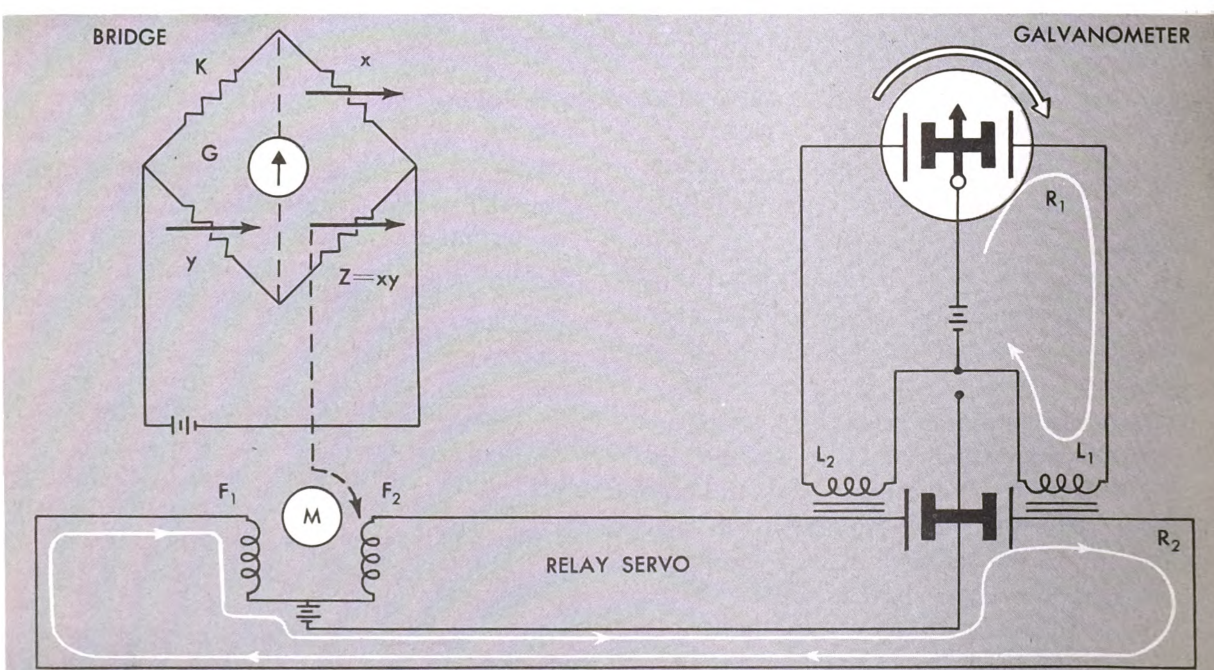
The Wheatstone Bridge Multiplier

The Wheatstone bridge circuit illustrated in the diagram may be used as a multiplier. The two variables to be multiplied appear as shaft rotations which position the variable resistances x and y in the bridge circuit. Resistance K is a fixed resistance and will be a constant scale factor in the circuit. The fourth leg of the bridge will represent the unknown quantity Z and will be used to balance the bridge. The rotation of the Z shaft will also represent the desired product of x and y. You will recall that when a bridge is balanced, no current flows through the galvanometer and the following ratios exist:

$$\frac{K}{x}=\frac{y}{Z},\ KZ=xy,\ \text{or}\ Z=\frac{xy}{K}$$

Let us assume that the rotations of x and y are produced automatically by servo rotors or by the precession of gyros, and that the bridge is maintained in balance with the use of the *relay servo* as illustrated. Any unbalance of the bridge circuit causes the galvanometer pointer to deflect either to the left or to the right, depending on the polarity of the unbalance. If the unbalance produces a deflection to the right, the galvanometer closes the relay circuit R_1. Current then flows through L_1, which closes the relay circuit R_2 thus allowing current to flow through F_1, the field winding of the servo motor which is mechanically linked to the Z resistor shaft. The Z resistor shaft is in turn rotated in the proper direction to rebalance the bridge. Hence the rotation of the Z shaft will be equal to the product of the rotation of the x and y shafts.

Upon re-examination of the operation of this multiplier, you will find that, if Z is introduced as a dividend and X is fed in as the divisor, the output, y will be: $y=\frac{KZ}{x}$. Hence the equipment may perform the process of division as well as multiplication.

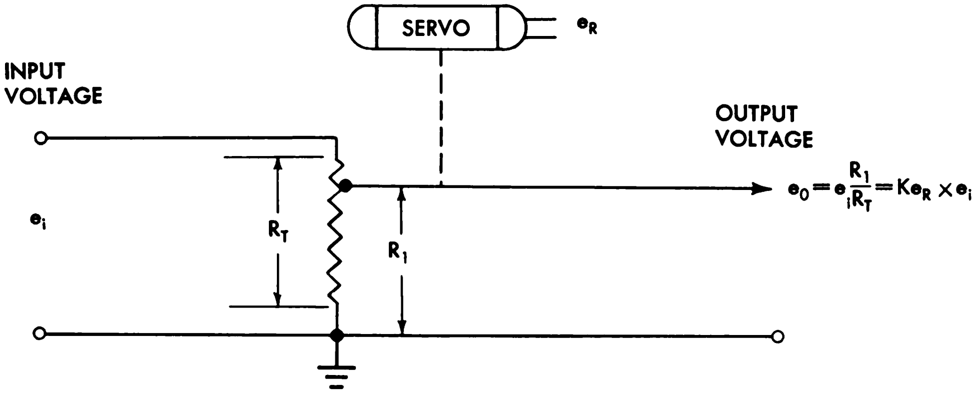


Wheatstone Bridge Multiplier

Potentiometer Multipliers

In AC and DC electronic computers, potentiometers are very often used for multiplication. They are simple, reliable, and capable of considerable precision.

In the potentiometer multiplier, the output voltage will be equal to the product of the two input voltages only when under no load or for as long as the measuring device to be supplied by e_o does not draw appreciable



A Servo Driven Potentiometer Multiplier

In the illustrated multiplication circuit, the output voltage e_o is the product of two variables present. One of these variables e_i is shown as the applied voltage to the input end of the linear potentiometer. The other variable e_R appears as a shaft rotation which picks off R_1, a portion of the total resistance R_T. With a linear potentiometer, the resistance R_1 will be directly proportional to the degree of the potentiometer shaft rotation. Also, the amount of shaft rotation may be made proportional to the input error signal designated as e_R. For example, at the setting indicated by R_1 in the illustration the proportion of input voltage e_i present at the output terminal e_o is:

$$\frac{e_o}{e_i} = \frac{R_1}{R_T}$$

And it follows that the open circuit voltage e_o at the potentiometer pickoff is equal to:

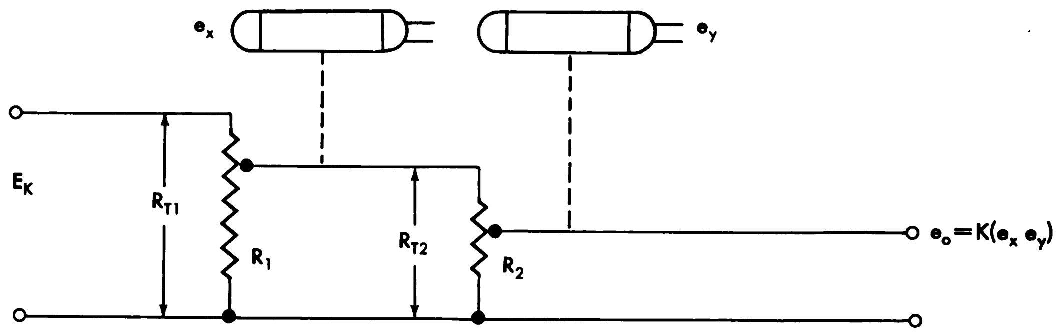
$$e_o = \frac{R_1}{R_T}\,(e_i) = Ke_R \quad e_i$$

$$\textit{where}\ \frac{R_1}{R_T} = Ke_R$$

current from the potentiometer (such as a high input impedance vacuum tube circuit). On the other hand, if the output voltage e_o is fed to a low impedance load, it would deviate from the proportionality of the servo shaft rotation. The amount of deviation will, of course, depend upon the degree of loading. As was pointed out in the section on mixers, attenuating or scale-factor resistors may be inserted to compensate for potentiometer loading.

CASCADING SERVO DRIVEN POTENTIOMETERS. In the example just discussed, one variable appeared as a voltage and the other appeared as a shaft rotation. Frequently both variables appear as shaft rotations. For example, two potentiometers may be cascaded to perform the multiplication, as shown in the next illustration.

Note that a constant reference voltage E_K is applied across the total resistance R_{T1} of the first potentiometer. The rotation of the first shaft (representing one of the variables) picks off a proportionate resistance R_1 of R_{T1}. The second shaft rotation (representing another variable) picks off a correspondingly proportional portion of R_{T2}



Cascaded Potentiometer Multiplier

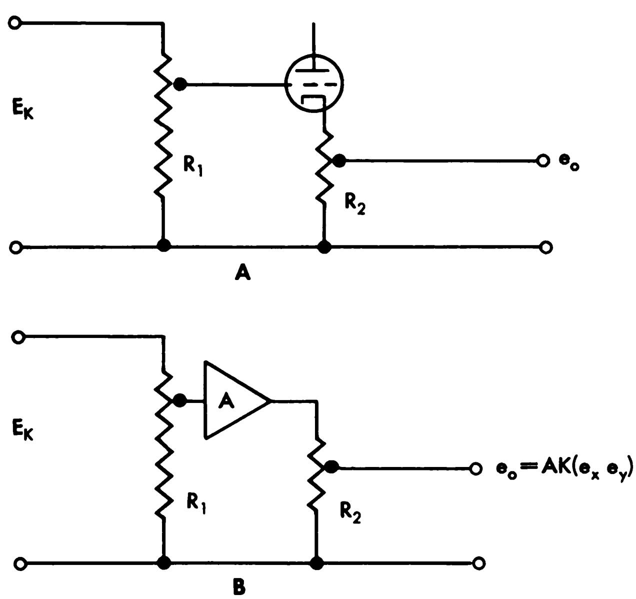
equal to R_2. If we *neglect* the loading of the second potentiometer on the first, the voltage e_1 at the pickoff point of the first potentiometer is equal to $E_K \frac{R_1}{R_{T1}}$. The out voltage e_o picked off of the second potentiometer is equal to $e_1 \frac{R_2}{R_{T2}}$. Upon substituting the value for e_1 in the second of the above equations we obtain the equation of the output voltage for the multiplier.

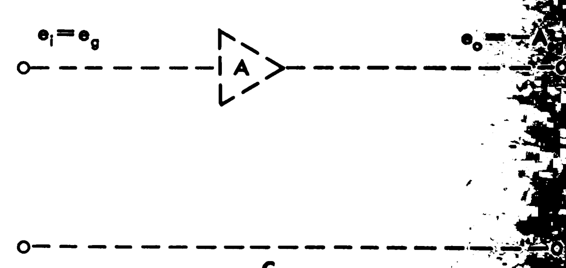
$$e_o = E_K\left(\frac{R_1}{R_{T1}}\right)\left(\frac{R_2}{R_{T2}}\right) = K(e_x\ e_y)$$

COMPENSATING FOR POTENTIOMETER LOADING. Because of the inherent loading effects when ganging potentiometers, the succeeding potentiometers should have a much larger resistance value than the preceding ones.

The loading effects produced by ganging potentiometers may be overcome by isolating one from the other with a one-way coupling device such as a buffer amplifier or cathode follower. You will recall that a cathode follower may be designed to have a high input impedance and a low output impedance.

The schematic at A and the symbolic representation at B illustrate two arrangements in which two variables may be multiplied together using two identical potentiometers, while at the same time overcoming the effects of loading with the use of buffer amplifiers. The conventional symbolic representation of an ideal DC amplifier with an infinite input impedance (without feedback) is indicated by the broken lines at C in the illustration. Ideally the output voltage

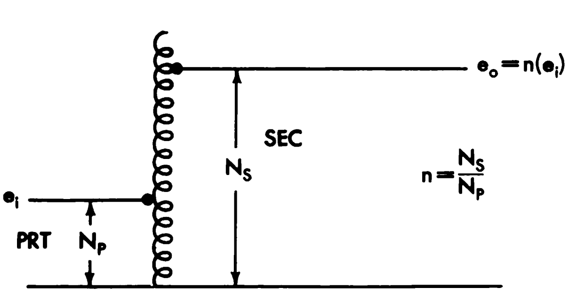




Potentiometer Multipliers with Buffer Amplifier

of the amplifier should be an exact magnified image of the input voltage e_i. The performance of the amplifier may be represented by the equation $e_o = -A\, e_i$ where A is the constant voltage gain. The ideal DC amplifier with negative feedback will be investigated more thoroughly after a brief discussion of the autotransformer used as a multiplier.

AUTOTRANSFORMER MULTIPLIER. The potentiometer multipliers so far described may be used with either direct current or with alternating current. If alternating current is used, multiplication of variables represented by shaft rotations may be accomplished with autotransformers. Such a transformer is illustrated schematically. Input

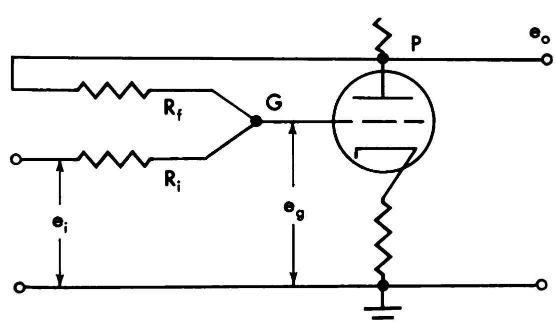


and pickoff brushes are allowed to slide on the exposed portion of the winding. When a portion of the winding is excited with an AC voltage source (such as e_i) the core becomes magnetized and a secondary voltage is induced across the entire winding. This secondary voltage will be equal to the product of e_i and n, where n is the transformer secondary to primary turns ratio. The secondary voltage also varies linearly with the pickoff brush position. A very desirable feature of the autotransformer multiplier is that autotransformers may be cascaded for further multiplication without the use of buffer elements to compensate for loading.

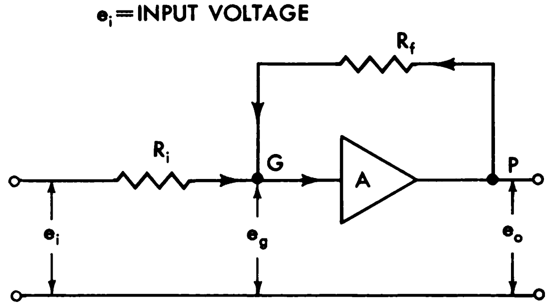
ELECTRONIC MULTIPLICATION AND DIVISION

Negative Feedback Operational Amplifier

The term *operational amplifier* denotes a high-gain DC amplifier used with appropriate input and feedback elements to perform various mathematical operations. You will recall that negative feedback is used to achieve linearity and to stabilize gain. Most amplifiers, except the cathode follower type, are arranged to produce an output that is A times the input voltage and of the opposite polarity. Among the characteristics of most DC feedback amplifiers, we find high input impedance, low output impedance, and very high-gain tubes. The conventional schematic **A** and the symbolic representation **B** of an operational DC amplifier (with negative feedback) are illustrated. The performance of the amplifier may be represented by the equations $e_o = A\, eg$ and $e_o = \frac{K}{\beta}\, ei$, where A is the constant voltage gain in the order of 100,000 or more and β is the feedback factor which is inversely proportional to the ratio of feedback to input impedance $\frac{(R_f)}{R_i}$.



R_i = INPUT IMPEDANCE (HIGH)
R_f = FEEDBACK IMPEDANCE
A e_i = INPUT VOLTAGE



e_g = GRID VOLTAGE
A = AMPLIFIER AND GAIN
B e_o = OUTPUT VOLTAGE

Sign Reversing and Multiplying by a Constant

One of the basic applications utilizing the high-gain DC amplifier in electronic analog computers is the sign-reversing amplifier. For example, let us assume the input impedance of the basic operational amplifier (shown above at B) to be infinite. Let us further assume that this is an ideal amplifier in which the net current flowing at the grid terminal is zero and all the current flowing through R_f also flows through R_i, making the current flow through R_i and R_f equal.

The equation for the current summation at the grid terminal G is then,

$$\frac{e_i - e_g}{R_i} + \frac{e_o - e_g}{R_f} = 0$$

And since $e_o = -Ae_g$ or $e_g = \frac{(-e_o)}{A}$, the above equation may be rewritten as,

$$\frac{e_i \frac{(e_o)}{A}}{R_i} + \frac{e_o \frac{(e_o)}{A}}{R_f} = 0$$

Because A is very large with respect to e_o, the terms $\frac{e_o}{A}$ and e_g may be neglected or considered to be zero. Hence, the above equations reduce to the following simple proportions:

$$\frac{e_i}{R_i} + \frac{e_o}{R_f} = 0; \text{ or } \frac{e_i}{R_i} = -\frac{e_o}{R_f}$$

Solving for e_o by multiplying through by R_f

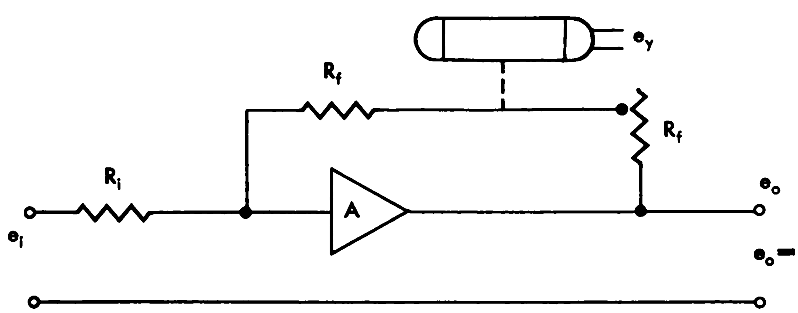
$$e_o = -e_i \frac{(R_f)}{R_i}$$

Thus the operational amplifier in this example serves as a multiplier by increasing or decreasing the input voltage e_i in accordance with the ratio of the feedback resistance to the input impedance, $\frac{R_f}{R_i}$, as well as changing its sign. If R_i is made equal to R_f, the above ratio becomes equal to unity and $e_o = -e_i$. In such an arrangement, the input voltage is simply multiplied by a constant, -1, to produce an output voltage equal in magnitude to the input voltage but of the opposite sign.

The ideal sign-reversing amplifier in the example and those to be discussed later are assumed to have an infinite gain A, yet with an output equal in magnitude to the input. In order to actually approach this theoretical situation, the input e_g to the grid must be equal to zero; therefore, the feedback must exactly nullify the voltage gain. That is, if $e_o = \frac{-A}{\beta} e_i$ and the ratio of voltage gain to feedback approaches unity, $e_o = -e_i$ very nearly. If it doesn't, the feedback may be changed by adjusting the scale factors—adjusting the ratio $\frac{R_f}{R_i}$.

Division by a Constant Coefficient or Variable

High-gain negative feedback amplifiers may be used for dividing variables by a constant coefficient. The diagram illustrates a circuit suitable for this operation. By means of the potentiometer in the feedback circuit, a certain constant percentage of the output



$$e_o = -Ae_i = \frac{-R_f}{R_i} e_i = \frac{k_i}{\beta} e_i = \frac{e_i}{Ke_y}$$

Division By a Constant Coefficient or a Variable

is fed back to the input. This feedback voltage is opposite in polarity to the input voltage, which reduces the gain of the amplifier by an amount proportional to the amount of voltage fed back. Here again $e_o = -\left(\frac{R_f}{R_i}\right)e_i$, and of course, the ratio of the feedback resistor to the input resistor could be made equal to one-half. Thus, if the output of the amplifier is reduced by one-half, it represents a division by two, $e_o = -\frac{e_i}{2}$. The amount of negative feedback which may be controlled by the setting of the potentiometer determines the constant by which the input voltage is divided.

The above circuit may also be used to divide one variable by another. The dividend (e_i), the voltage to be divided, is supplied to the input of the amplifier. The divisor voltage (e_y) is used to position the arm of the feedback potentiometer. Since the gain of a negative feedback amplifier is inversely proportional to the amount of feedback, the quantity $\frac{k_1}{\beta}$ may be substituted for A in the equation for output voltage ($e_o = -Ae_i$), giving $e_o = -\frac{k_1}{\beta}e_i$. And if feedback is made proportional to e_y, the feedback factor $\beta = k_2 e_y$. When substituting this quantity for β in the above equation, the equation for the output voltage becomes;

$$e_o = -\frac{k_1 e_i}{k_2 e_y} \text{ or } e_o = \frac{e_i}{Ke_y}$$

giving the result of dividing e_i *by* e_y.

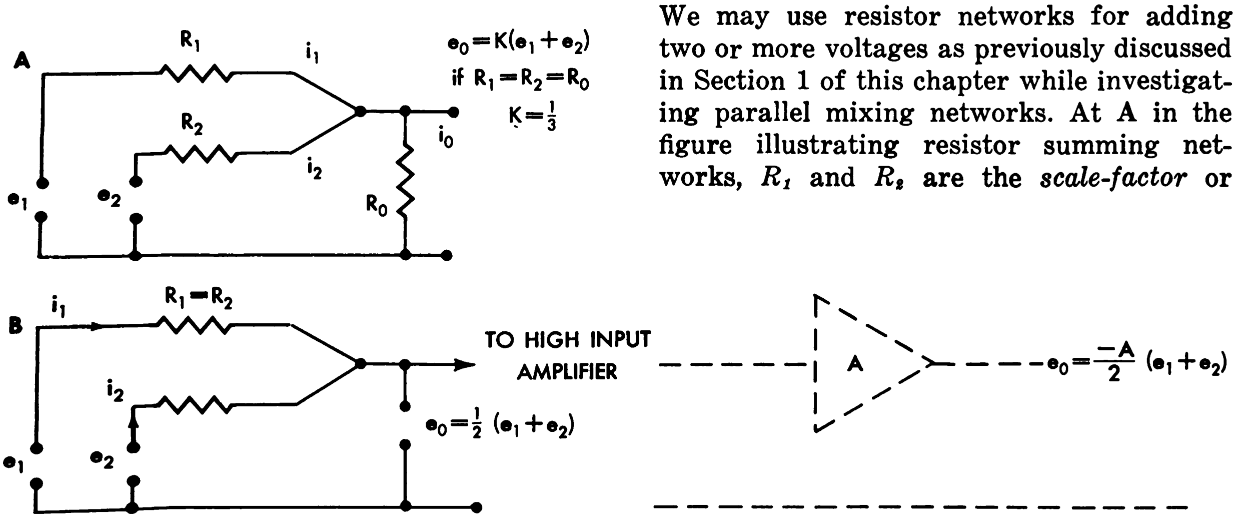
Although it is sometimes necessary to use dividers in analog computing networks, they are avoided whenever possible. It is impossible to design a divider that will function correctly as the divisor approaches or passes through zero. It can be seen from the above equation that, if e_y approaches zero, the quotient e_o must approach infinity. Hence, any computer that uses a divider should be arranged so as not to demand division by small numbers to prevent overloading the amplifiers used. It is also desirable to keep the range of dividers as small as possible.

Another fundamental circuit used to accomplish the process of division utilizes a multiplier unit as well as a high-gain negative feedback summation amplifier. For obvious reasons this circuit will be investigated after discussing summing amplifier networks.

ELECTRONIC SUMMING NETWORKS

Summing Resistors

We may use resistor networks for adding two or more voltages as previously discussed in Section 1 of this chapter while investigating parallel mixing networks. At A in the figure illustrating resistor summing networks, R_1 and R_2 are the *scale-factor* or



Resistor Summing Networks

summing resistors. The voltage division between the output resistor R_o and each summing resistor results in an attenuation of each input voltage applied to the circuit. If the summing resistors are of equal value, the output voltage will be proportional to the sum of the two input voltages. That is, $e_o = K(e_1 + e_2)$, where the value of K is determined by the ratio of the output resistor to the summing resistors.

For example, in the above circuit, the current flowing through the output resistor R_o is equal to the sum of the currents flowing through R_1 and R_2 — that is,

$$i_o = i_1 + i_2$$

$$or: \frac{e_o}{R_o} = \frac{e_1 - e_o}{R_1} + \frac{e_2 - e_o}{R_2}$$

If $R_1 = R_2 = R_o$, the above equation may be rewritten as;

$$\frac{e_o}{R} = \frac{e_1 - e_o}{R} + \frac{e_2 - e_o}{R}$$

Clearing fractions by multiplying through by R,

$$e_o = e_1 - e_o + e_2 - e_o$$

Collecting like terms,

$$3\ e_o = e_1 + e_2$$

$$e_o = \frac{e_1 + e_2}{3}$$

Under the condition of no load, or considering the load to be the infinite input impedance to an ideal amplifier without feedback (as shown by the broken lines at **B**), the current sum at the output terminal of the summing network (input terminal of the amplifier) is:

$$\frac{e_1 - e_o}{R_1} + \frac{e_2 - e_o}{R_2} = 0$$

Solving for e_o,

$$e_o = \frac{R_2}{R_1 + R_2}\ e_1 + \frac{R_1}{R_1 + R_2}\ e_2$$

Note that the summation includes the scale factors (coefficients) of e_1 and e_2, which may be adjusted by assigning the appropriate values to R_1 and R_2. If $R_1 = R_2$, you will find, when substituting, that $e_o = \frac{e_1 + e_2}{2}$.

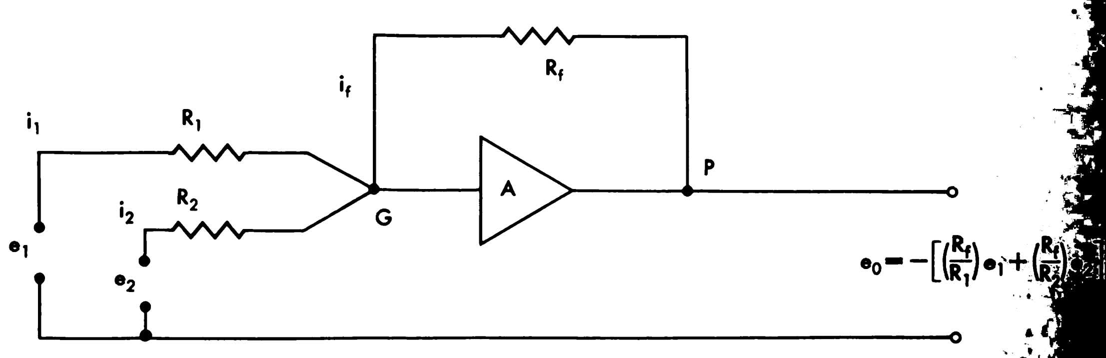
The summing networks shown utilized only two input resistors; however, they may have any number within reason. In general then, $e_o = \frac{e_1 + e_2 + \ldots e_n}{n}$ which is the output without amplification. This shows us that, without amplification, e_o is the arithmetic mean of all voltages, the identical resistors (the usual case) serving as an averaging circuit. If the amplifier without feedback is included, the output equations are multiplied by the gain A as indicated in the illustration.

Summing Amplifiers With Negative Feedback

Normally a high-gain negative feedback amplifier is utilized as a buffer stage to isolate the actual load from the summing network as shown in the accompanying illustration. Note that this *Summing Amplifier* is similar to the basic sign-changing amplifier previously discussed. If we assume that the net current at the grid is again zero, the current summation at the grid terminal may be represented by the following equations:

$$i_1 + i_2 + i_f = 0$$

$$\frac{e_1 - e_g}{R_1} + \frac{e_2 - e_g}{R_2} + \frac{e_o - e_g}{R_f} = 0$$



Summing Amplifier

And since $e_g = \frac{(-e_o)}{A}$ where A is very large with respect to e_o, the term e_g in the above equation may be neglected or considered to be zero. Hence, the current summation equation reduces to the following simple proportions:

$$\frac{e_1}{R_1} + \frac{e_2}{R_2} + \frac{e_o}{R_f} = 0$$

or
$$\frac{e_1}{R_1} + \frac{e_2}{R_2} = -\frac{e_o}{R_f}$$

Solving for e_o by multiplying through by R_f,

$$e_o = \left[\left(\frac{R_f}{R_1}\right)e_1 + \left(\frac{R_f}{R_2}\right)e_2\right]$$

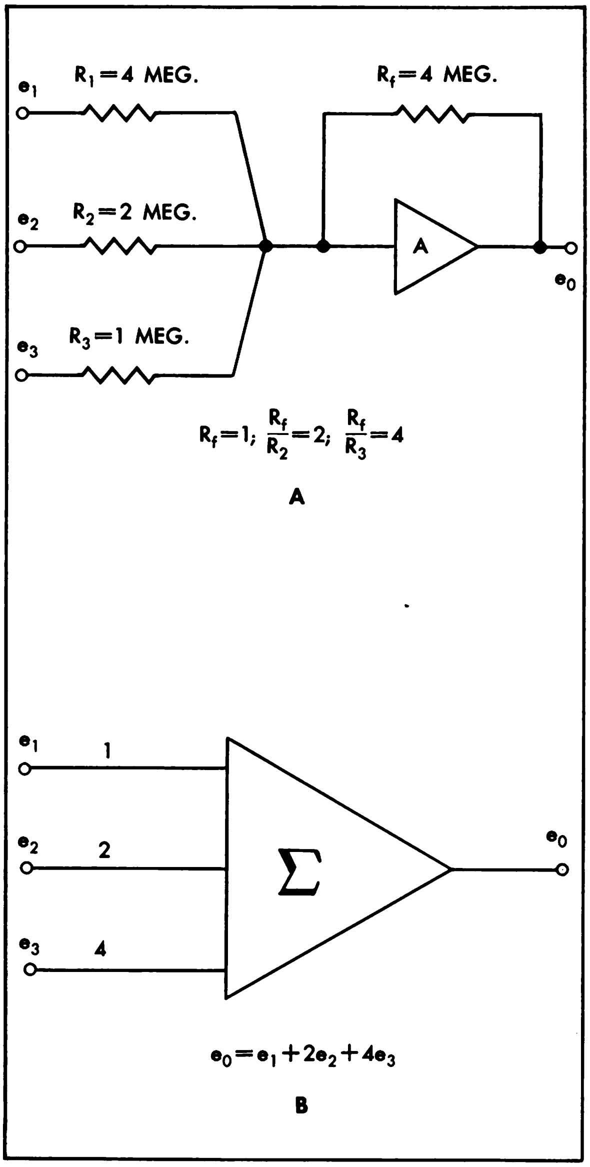
Note that here again the summation includes the scale factors (coefficients of e_1 and e_2) which may be adjusted by the appropriate assignment of values to R_1, R_2, and R_f. If $R_1 = R_2 = R_f$, the ratios of the feedback resistor to the summing resistors (coefficients of e_1 and e_2) become equal to unity and the above equation reduces to $e_o = -(e_1 + e_2)$.

Hence, the summing amplifier with equal input and feedback resistors produces an output voltage that is equal to the sum of the input voltage multiplied by a constant, -1. If the values of the summing and feedback resistors are unequal, each input voltage is multiplied by its respective scale factor before summation takes place.

SUMMATION AMPLIFIER REPRESENTATION. Many of the diagrams of computing networks containing summing amplifiers simply indicate the scale factors at each input terminal without showing the summing and feedback resistors in the block diagram. For example, consider the summing amplifier and its accompanying equivalent circuit illustrated in the next diagram. The actual network shown at **A** consists of three summing resistors. With these values as indicated, the coefficients of e_1, e_2, and e_3 are respectively as follows:

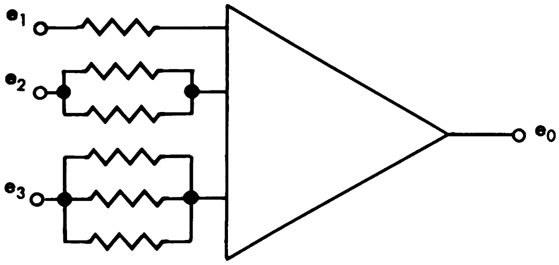
$$\frac{R_f}{R_1} = 1;\ \frac{R_f}{R_2} = 2;\ and\ \frac{R_f}{R_3} = 4$$

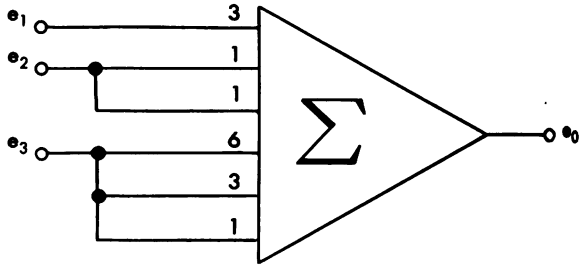
The equivalent circuit at **B** shows a commonly used block diagram symbol with the three input terminals corresponding to the above coefficients. The Σ (summation sign) signifies that the amplifier is used as a summing device to add the input voltages representing variables.



Summing Amplifier and Equivalent Circuit

At times it is desirable to have a voltage representing a variable added to itself or to produce the summation of variables that are first multiplied by several constant coefficients. This is often accomplished by feeding the same input voltage to two or more input terminals of the summing amplifier — as is shown in the diagram illustrating the *Parallel Input Summing Amplifier*.





$e_0 = 3e_1 + 2e_2 + 10e_3$

Parallel Input Summing Amplifier

It is also often desirable to accomplish the summation of variables which are first multiplied by other variables. In such cases you may find the variables introduced by allowing them to be fed into multiplier boxes which are located at points in the input leads of a high-gain amplifier. Two typical circuits of this type — indicating the summation of the product of two or more variables — are illustrated at **A** and **B** in the next diagram.

The Summation of Products Circuit composed of the summation amplifier and the multiplier unit (illustrated at **B**) may be used to accomplish the process of division. Let us assume, for example, that the circuit is used to determine the quotient of the two input variables e_i and e_y. The variable e_i is fed as the dividend directly to one terminal of the summation amplifier. The divisor voltage e_y is fed first to the multiplier unit to be multiplied by Ke_o, where Ke_o is a portion of the output voltage (feedback) which is also fed to the multiplier unit. The output of the multiplier — the product $(e_y\, e_o)K$ — and the dividend e_i are then summed in the amplifier. The equation representing the resulting output of the summation amplifier may be expressed as:

$$e_o = -A[(e_y\, e_o)\, K + e_i]$$
$$e_o = -A\,(e_y\, e_o)\, K - Ae_i$$
$$e_o + A(e_y\, e_o)\, K = -A\, e_i$$

Since e_o is very small as compared to $A(e_y\, e_o)\, K$, the first term in the last equation above may be neglected. Hence;

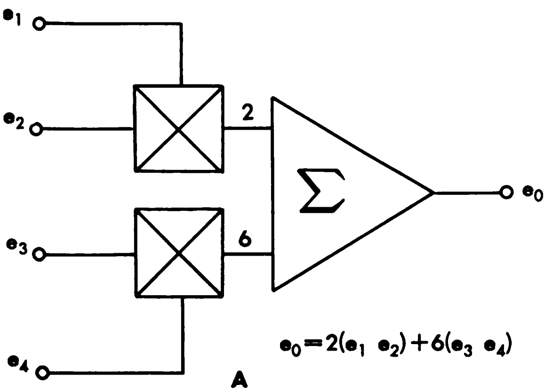
$$A(e_y\, e_o)\, K = -Ae_i$$

Solving for e_o,

$$e_o = \frac{-A\, e_i}{AK\, e_y}$$

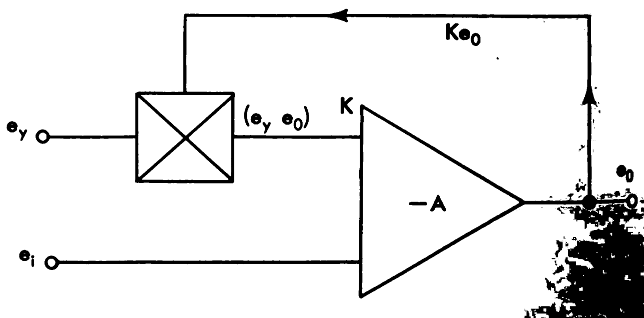
Thus the output of the amplifier after feedback becomes the quotient of e_i divided by e_y.

$$e_o = \frac{-e_i}{Ke_y}$$



$e_0 = 2(e_1\, e_2) + 6(e_3\, e_4)$

A



$e_0 = -A[(e_y\, e_0)\, K + e_i]$

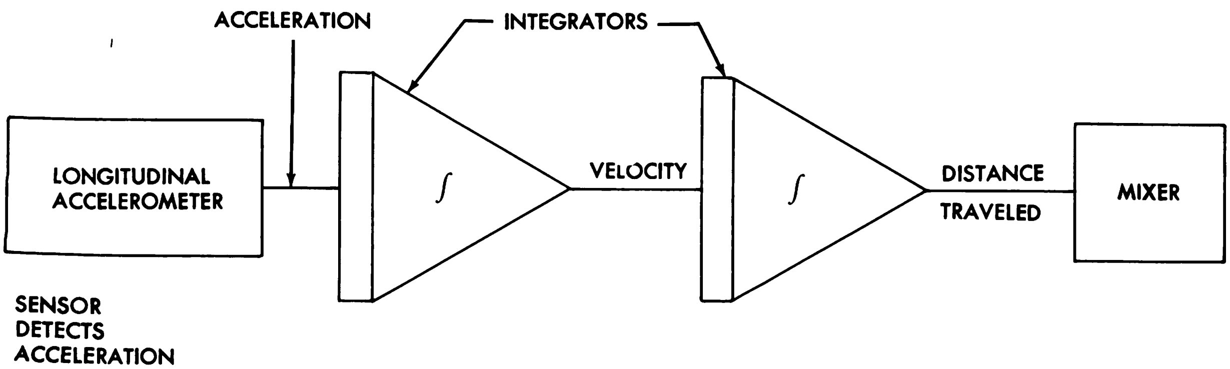
B

Indicating the Summation of Products

SECTION C, INTEGRATOR COMPONENTS OF COMPUTERS

An integrator is that part of a computer system which, in effect, performs a mathematical operation on an input signal. Simply stated, the integral of a constant signal is proportional to the amplitude *multiplied* by the time the signal exists. For example, suppose the output of an electronic integrator is four volts, and this output was produced by a constant input signal which lasted for one second. If the same input signal had lasted for only one-half second, the output would be two volts. Furthermore, if the signal were twice as large and lasted for one second, the output would be eight volts. Thus, an integrator is a device for considering the *time* that an input signal exists and for changing the input to a form representing *new information.*

Normally the actual detected error of acceleration is not constant. Instead, the amplitude and sense of the error changes according to the attitude of the interceptor aircraft or missile. Regardless of this fact, the correct integration will be proportional to the product of the operating time and the average error prevailing during this time interval. If the sense of the error should change during the integration period, the signal of opposite sense would decrease the final output of the integrator. The integrator can be considered to be a continuous computer since it is always producing a voltage which is proportional to the *product* of the *average input* and the time.



Integrators Contained in the Longitudinal Channel of an Inertial Guidance Computer

INERTIAL GUIDANCE INTEGRATORS

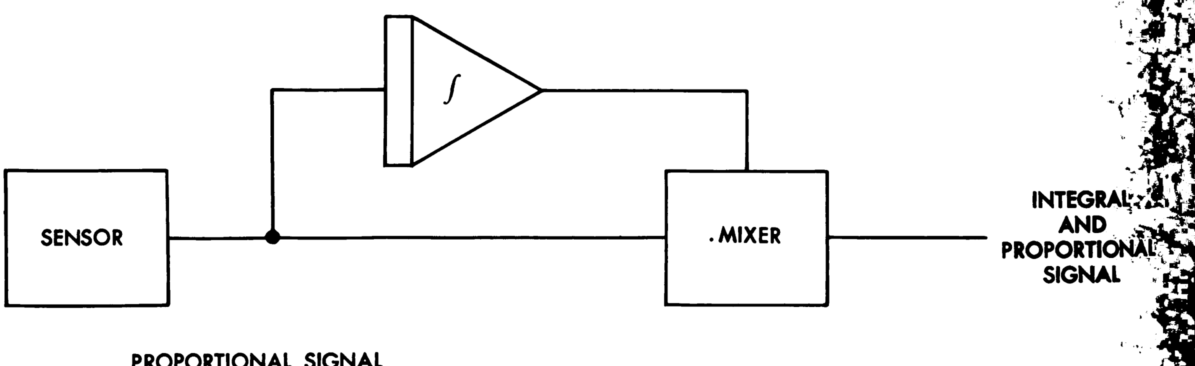
The integrators contained in the longitudinal channel of an inertial guidance computer (illustrated) are required to determine time, velocity, and distance from a detected error of *acceleration.* In comparison you will recall that the output of a sensor which provides continuous control is normally instantaneously proportional to some attitude error. This proportional signal provides one fundamental means of continuous control. *Integral* and rate or *differential control* are two other methods for providing such control.

It may be stated that the integration of an error with respect to time represents the summation or accumulation of errors over a specified period of time. Actually, one definition by Webster for the term integrate is "to unite (parts or elements) so as to form a whole; also, to unite (a part or element) with something else, esp. something more inclusive; to give the sum or total of."

There are times in intercepter aircraft and missile systems when a signal proportional to an accumulation of errors over a period of time is very important. For example, in an inertial guidance system an integrator control signal is combined with a proportional signal from a sensor by means of a mixer as illustrated. The integrator signal will either support the error signal or subtract from it, depending on the purpose.

Subtracting Signal

An integrator signal subtracts from the proportional signal to eliminate any undesirable signal component. The general connections are the same as shown in the previous figure except that the sense of the integrator output is reversed. When the proportional signal and integral signals counteract, the result is to reduce or eliminate an undesirable steady component of the proportional signal.



Combining a Proportional Signal with Its Integral Component

Supporting Signal

An integrator signal is used to support the proportional error signal to insure that sufficient correction will always be made by a system. The degree of control that a pure proportional signal can exert in a system is limited. There may be a time when proportional control alone is not sufficient to overcome a strong, steady force (say, a crosswind) causing a deviation in attitude. In such a case the proportional error signal will have a steady component which will be fed to both the integrator and the mixer. Since the error will be of one sense, the integrator output will increase with time and augment the proportional signal until correction ultimately takes place. The integrator output continues to supply the necessary correction signal component to overcome the relatively constant influence of error due to a steady crosswind.

Such a need may exist in self-balancing an amplifier output. The signal caused by unbalance will appear as a steady component of the total output and will be opposed by the integral of that output. Similarly, any unbalance which may arise from a sensor, such as synchro misalignment or gyro unbalance, may be counteracted.

Normally, the pickoff which provides the output is zeroed before the beginning of integration. Sometimes, however, an initial reference other than zero (or null as it is often called) is required. The integrator would then begin with a predetermined constant input and output. Such a beginning signal is important because it represents a condition which should exist at the beginning of integration and is called the *initial condition.* Applications of supporting and subtracting integral control will be investigated

further during the discussion of an inertial guidance system in a subsequent chapter.

DIFFERENTIAL ANALYZER

A differential analyzer is an analog device which may be used for simulating a dynamic system involving time, motion, and distance. It is a device utilized to accomplish the solution of rate problems and differential equations where the initial conditions may be predetermined and fed into the machine. For our purpose, the solution commonly takes the form of an indication as to how some variable such as distance or velocity changes with respect to time in response to some impressed force such as a detected error of acceleration.

In order to understand how velocity and distance are determined from acceleration, new mathematical equations must be set up which show the relationship of one factor in all its variations to each of the other factors and their variations. These equations are called differential equations and may be expressed as follows:

1. *Acceleration* is the rate at which an object is changing velocity with respect to time. Expressed as a differential equation

$$a = \frac{dv}{dt}$$

where ($dv = \Delta v$) is a small differential change in velocity with respect to a correspondingly small differential change in time.

2. *Velocity* is the rate of change of distance with respect to time. Expressed in differential form

$$v = \frac{ds}{dt}$$

Expressed in integral form,

$$v = \int adt = \Sigma adt$$

which states that velocity is equal to the integral of acceleration with respect to time.

3. *Distance* is the rate of change of velocity with respect to time. Expressed in differential form,

$$s = \frac{dv}{dt}$$

Expressed in integral form,

$$s = \int vdt = \Sigma vdt$$

which states that distance is the integral of velocity with respect to time.

The solution of these equations involves either the mathematical process of *integration* or *differentiation*. A few of the basic mechanical and electronic methods of accomplishing the process of integration will be discussed in subsequent portions of this section.

Rate or differential control and the methods of accomplishing the process of differentiation will be investigated in the next section of this chapter. The principles involved in the solution of rate problems and differential equations will now be discussed.

A Basic Rate Problem

Of course, there is no advantage in solving simple rate problems with the use of a differential analyzer. However, a basic rate problem will aid us in understanding the principles involved.

For example, if we know the constant velocity of a missile and the time the missile travels at this velocity, we can determine the distance of travel. That is, the total distance traveled will be equal to the product of the constant velocity and the total time of travel, which may be expressed in equation form:

$$s = vt$$

Solving for v,

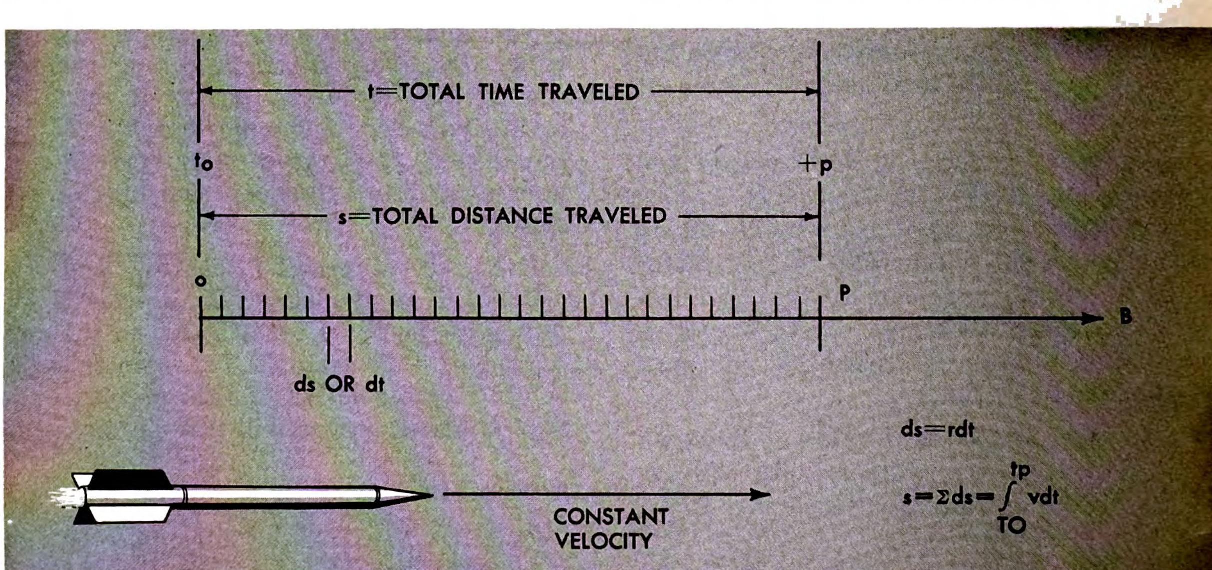
$$v = \frac{s}{t}$$

Hence, velocity or speed of motion may be

defined as *the time rate of change of distance* which is equal to the space passed over divided by the time required to pass over it.

It may be helpful to visualize the above familiar facts and those which follow by referring to the graph illustrating simple linear motion. In the graph, it is assumed that motion takes place along the straight line *OB* in the direction *O* to *B* as indicated by the arrow. Point *p* represents the position of a missile in motion at any particular instant. The space traveled by the missile from the reference *O* to point *p* is designated as *s*.

tionally regarded as a kinematic differential equation of motion. Let us now assume that it is desired to solve this equation for the total distance traveled during the time interval t_o till t_p. If the solution were to be obtained by mathematical integration or by means of an analog computer, it would give the numerical distance *s* as a function of time *t* at any desired time t_p. In other words it will show how the position of the missile varies with time. In either case the solution involves the process of integration; that is, it involves taking the integral (accomplishing the summation or accumulation) of all



Illustrating Simple Linear Motion

This distance *s* may be considered to be composed of a large number of small differential fractions (Δs or ds) of the total distance. The total time of travel may also be considered to consist of a large number of small differential increments of time (Δt or dt). If it is desired to determine a differential distance ds during a corresponding increment of time dt, we could multiply the differential increment of time by the given constant velocity.

$$ds = vdt$$

The equation directly above is conven-

the differential distances (ds) prevailing during the specified time interval t_o till t_p. Where each differential increment of distance is in turn determined by multiplying each differential increment of time (dt) by the constant velocity that prevailed during the specified time interval t_o till t_p.

Mathematically integrating the equation

$$ds = vdt$$

we write, by taking the integral of both sides:

$$\int_{t_o}^{t_p} ds = \int_{t_o}^{t_p} vdt$$

or

$$\sum_{t_o}^{t_p} ds = \sum_{t_o}^{t_p} vdt$$

where the integral sign $\int_{t_o}^{t_p}$ or the summation sign $\sum_{t_o}^{t_p}$ specifies that all the differential quantities following the sign, which are present during the specified time interval t_o till t_p, must be added or accumulated before being multiplied by the constant velocity. The summation of all the differential distances (ds) present from t_o till t_p is equal to s; therefore, the above equations reduce to

$$s = \int_{t_o}^{t_p} vdt$$

which again states that the total distance traveled is equal to the integral of the velocity with respect to time.

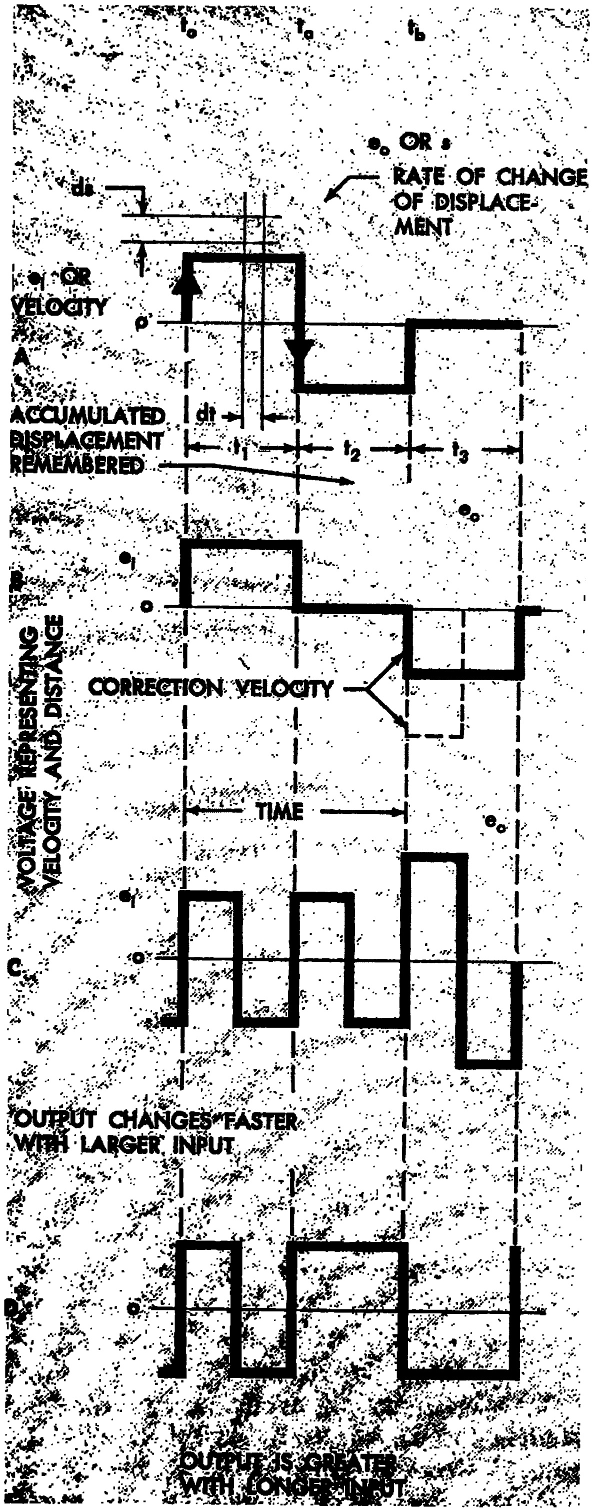
The Reaction of an Ideal Integrator

Some types of integrators are more accurate than others because of their principles of operation. There are several problems involved in producing an integrator which will deliver an output ideally proportional to the integral of the input. First, the integrator must be able to react linearly to any signal level which may be applied to it. That is, if the input signal becomes doubled, the output of the integrator should increase at twice the previous rate. Secondly, this reaction should occur rapidly. Thirdly, the integrator must have the ability to *remember*. For example, if a certain signal is developed at the integrator output and there is no further input signal, the previous output should remain at the output for a time —until correction can be effected.

If you carefully examine the waveforms *Illustrating the Reaction of an Ideal Integrator to Various Input Signals,* you will see how sense of the integrated output depends on the average input. For example, at **A** in the illustration, the input voltage (representing velocity) suddenly reaches a certain value above the zero reference line. At this time t_o, no time has elapsed; so the integrated output voltage, representing position displacement or distance off course, is zero at this instant. However, the input voltage (velocity) remains constant during the time interval designated at t_1; hence the output voltage, representing distance traveled off course, increases at a linear rate during this time interval. The total distance or deviation off course accumulated within this time interval is indicated as s and is represented by the ultimate magnitude of e_o. At time t_a, the input voltage suddenly reverses sense with the same magnitude and for the same period of time, representing a constant velocity which is now in the opposite direction. Hence, the integrated output returns to its initial value. In other words, when the deviation occurs in the opposite direction with the same constant velocity and for the same length of time, the ultimate displacement is reduced to zero—at the same rate. This shows that the integral of a square wave is a triangular wave.

Now notice (at **B** in the illustration) the voltage representing velocity suddenly decreases to zero at time t_a and remains at this reference level for the time interval designated as t_2. This causes the integrated output to suddenly stop rising, level off, and remain constant for the duration of this time interval. Hence, if the displacement velocity suddenly becomes zero after a certain amount of distance off course has been accumulated, the magnitude of this displacement must remain fixed. Furthermore, the magnitude of this displacement remains fixed (is remembered) until a correction signal, producing a velocity in the opposite direction, is applied to again reduce the accumulated displacement to zero.

If the correction velocity applied at time t_b has the same magnitude and time duration as in the previous example (represented here by the reversed input voltage during interval t_3), the accumulated displacement is again reduced to zero at the same rate.



Illustrating the Reaction of an Ideal Integrator to Various Input Signals

However, if the correction velocity has a greater magnitude for a shorter time duration, the integrated displacement is reduced to zero at a faster rate as indicated by the voltage waveforms with broken lines.

The waveforms at C in the illustration show how the output changes faster when the input becomes larger. The waveforms at D illustrate how the output becomes greater if the input is applied for a longer period of time.

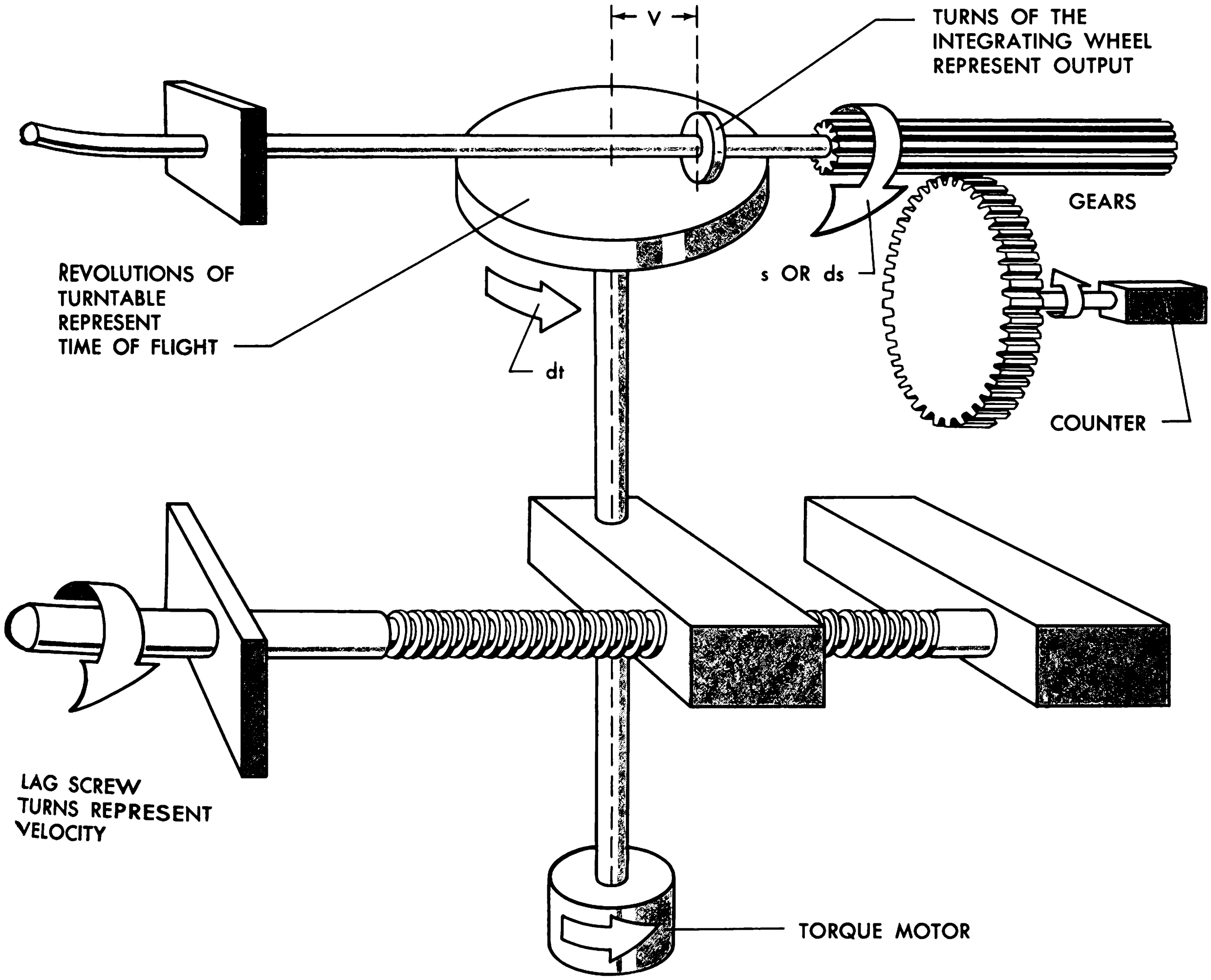
MECHANICAL INTEGRATORS

Any device capable of receiving and storing physical quantities and indicating the quantity stored may be used as an integrator. One example of such a *mechanical* device is the *Disc and Wheel Integrator* illustrated. In this type integrator, the rotations of the small integrating wheel are accumulated to represent a distance or displacement, and the accumulated revolutions of the wheel representing displacement may be recorded by a number in a mechanical counter as shown.

Disc and Wheel Integrator

Note that the disc and wheel integrator consists of three principal parts: a *disc* which is frequently called a *turn table*, a small *integrating wheel*, and a *lag screw* to position the integrating wheel. The turn table rotates (*t* or *dt* turns) horizontally on a vertical shaft. The integrating wheel rests upon the turn table and rotates (*s* or *ds* turns) on its shaft which is parallel to the surface of the turn table. The lag screw is threaded through the supporting blocks for the shaft upon which the turn table is mounted. Thus, as the screw is turned (*v* turns), the distance *V* between the center of the turn table and the edge of the integrating wheel will also vary.

The amount that the wheel turns increases as the distance *V* between it and the center of the turn table is increased. If the wheel is brought closer to the center of the turn



Disc and Wheel Integrator

table, the amount of turning decreases. Therefore, the total amount the wheel turns is equal to the sum of all the small amounts of turning multiplied by the distance V between the center of the disc and the edge of the wheel at the instant of turning. In this manner, the device expresses integration because the total amount that the integrating wheel turns equals the integral of the distance from the center of the disc to the edge of the wheel with respect to the revolutions of the turn table.

Now let us assume that the turns v of the lag screw, which determines the distance V, represents a measure of the *velocity* of an interceptor aircraft or missile. Let us further assume that the lag screw remains fixed for a very small increment of time, thus maintaining a fixed distance V and representing a *constant velocity* input to the integrator during this increment of time. Let us also assume that the turns (t or dt) of the turn table represent the time of flight—the differential input variable to the integrator. The turns (s or ds) of the integrating wheel will then represent the output, which is computed distance.

Considering a very small increment of time, the turn table rotates through a differential fraction of a turn dt which causes

the integrating wheel to turn through a correspondingly small differential portion of a turn ds. Hence,

$$ds = Vdt$$

During a long time interval (time of flight, say, of t_o until t_p), the turn table will turn through a larger portion of one complete revolution or a finite number of complete revolutions. The distance designated as V, representing the velocity, may remain fixed, representing a constant velocity; or it may vary from the center to the outer edge of the turn table, representing a variable velocity. In either case the total number of turns registered by the integrating wheel will be

$$s = \int_{t_o}^{t_p} Vdt$$

Thus the device may be used to mechanically integrate velocity with respect to time, giving the distance as shown by the above equation.

Integrating Gyro

Precession of a gyroscope is often used to accomplish the process of integration, particularly in cases where the *stabilizing properties of a gyro are also desired.* One of the requisites of an ideal guidance system specifies that the system be invulnerable to enemy jamming. To date, only two of the many proposed guidance systems, the celestial system and the inertial system, meet this requirement. Both of these guidance systems depend on gyros for their basic reference. The celestial system must have a gyro-stabilized platform for its telescopes, and the inertial system must have a stabilized platform for its accelerometers. Since the main cause of random drift in gyros is gimbal friction, the problem of improving gyro stability has been approached with a view to reducing this friction.

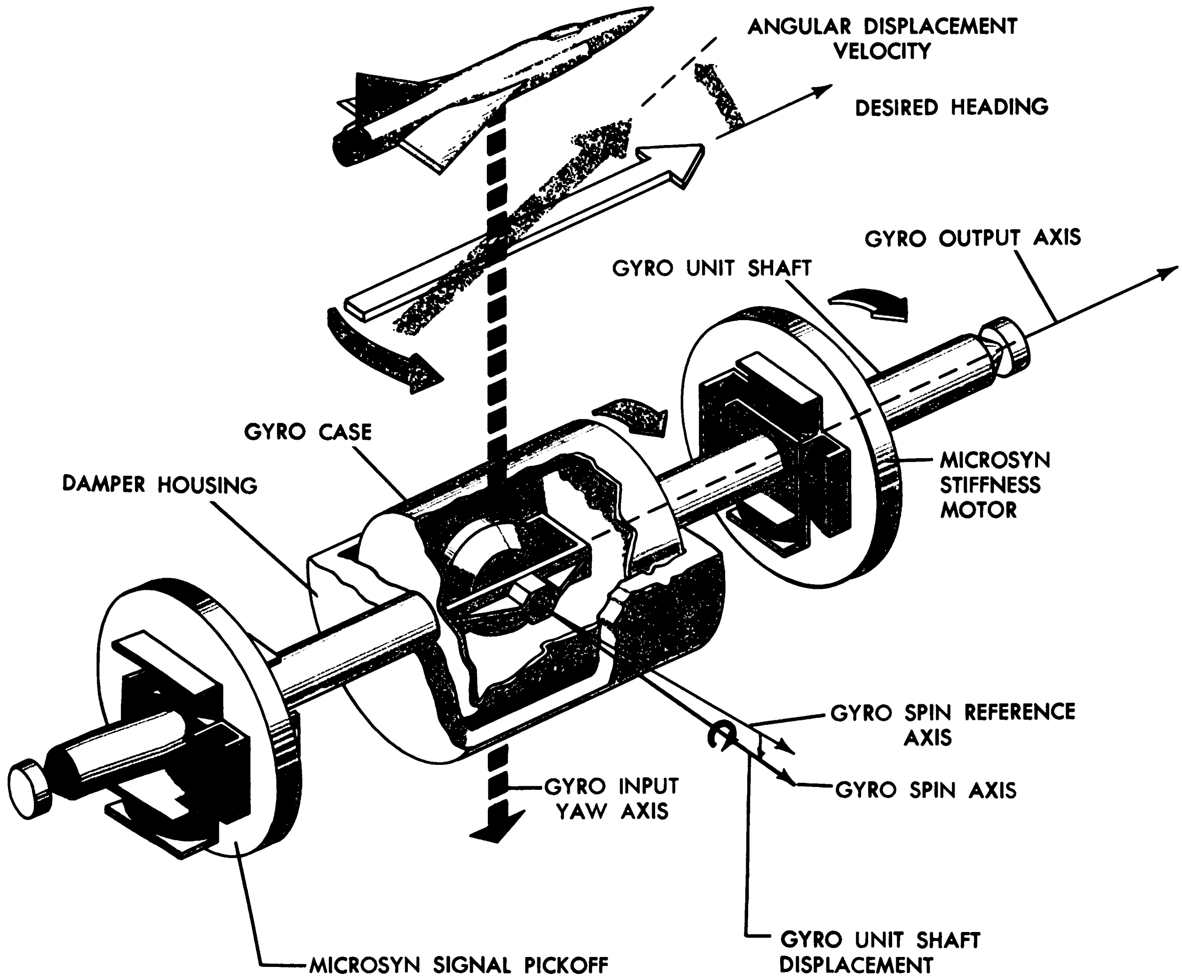
VISCOUS-DAMPED INTEGRATING GYRO. A viscous-damped integrating gyro is a good example of the progress that has been made toward the development of more accurate gyros for integral (attitude and guidance) control in interceptor aircraft and missiles.

The illustration shows a cutaway view of a *Viscous-Damped Integrating Gyro Unit.* The gyro rotor is contained within the gyro case. The space between the gyro case and the damper housing is filled with viscous damping fluid of high specific gravity. Because of the high specific gravity of the fluid, it serves to float the gyro case and gimbal shaft, thus reducing gimbal bearing friction. In this way, the random drift due to friction is greatly reduced. Because the gyro is supported by the viscous damping fluid, the unit is relatively unaffected by acceleration and other outside forces.

The Integrating Function is produced as the gyro case is allowed to precess through the restrictive force of the viscous damping fluid. Note that the gyro gimbal and case are free to rotate, with respect to the damper housing, about a single axis. This axis is called the output axis and is shown to be perpendicular to the rotor spin axis. The third axis, the input axis, is mutually perpendicular to both the spin axis and the output axis.

The gyro unit is assumed to be mounted in an interceptor so that its input axis is parallel to the vertical yaw axis and its output axis is parallel to the line of flight. If the interceptor begins to deviate off course (in yaw), the gyro case is subjected to an angular velocity with a component of displacement torque about the input axis; hence, a precessional torque develops about the output axis. You will recall that, when the rotor of a gyro is resolving at a constant speed, the rate of precession—and hence the precessional torque—is directly proportional to the displacement torque applied about an axis at right angles to the axis of precession.

The displacement torque applied at the input axis, in our example, is produced by the angular velocity or rate of change of interceptor off-course displacement. Hence, the integrated output—the amount of precession—is equal to the rate of precession multiplied



Viscous-Damped Integrating Gyro

by the time in which precession takes place. In other words, the amount of precession that is accumulated over a specified interval of time becomes the integrated output. This accumulated amount of precession is proportional to the accumulated amount of off-course displacement which is in turn proportional to the product of time and the angular off-course displacement velocity prevailing during this time.

The Microsyn Signal Pickoff, shown mounted on the left side of the gyro output shaft, produces a signal which is proportional to the integrated amount of precession. This signal is then fed to a command circuit to initiate proper corrective action. The command circuit may be similar to the one discussed in the previous chapter when investigating the application of gyros in servo loops.

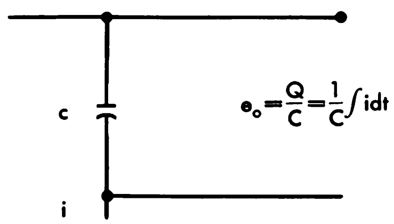
The Microsyn Stiffness Motor, shown mounted on the right side of the gyro output shaft, produces a torque that may either support or oppose the precessional torque. For example, it may be desirable to have the interceptor make a coordinated turn or change heading at some point in space. The required angular velocity to produce the new heading may or may not be in the same direction as the error angular displacement velocity. Hence, the actual torque applied to the output shaft and the pickoff signal commanding

system action will be proportional to the algebraic difference between the desired angular velocity of the interceptor and its actual angular velocity about the yaw axis.

The integrating gyro discussed had only one degree of precessional freedom; however, a gyro can be made to precess about two axes simultaneously and thereby integrate two rate components of an angular movement of a point in space. You may find this technique used in various fire control systems to continuously generate the bearing and elevation of an aircraft target on the basis of its angular rates—and at the same time stabilize the sight, gun, or missile launcher against rolling motion or maneuvering of, say, a fighter-interceptor aircraft.

ELECTRONIC INTEGRATORS

You may recall from your knowledge of electrical circuit theory that a capacitor will store an electric charge ($Q = Ce_o$) and indicate the quantity stored by the voltage existing across its plates ($e_o = \frac{Q}{C}$). For example, consider the simple capacitor shown. For a



given capacitor C, the capacitor voltage will depend directly on the coulombs of charge, Q, which is the balance of electrons on the two plates. While the output (indicating the quantity of charge stored) is a voltage e_o, the input is the charging current i which determines the rate at which electric charge enters the capacitor. Remember that the coulomb is a unit defining a quantity of electricity. A coulomb is the quantity of electrons moving past a point per unit time —the quantity of electrons transferred by a current of one ampere in one second. Hence, Q may be expressed as the product of the charging current and the time which this current flow exists—that is,

$$Q = i\,t$$

Substituting this quantity for Q in the equation for output voltage

$$e_o = \frac{Q}{C}$$

the equation for output voltage becomes:

$$e_o = \frac{1}{C}\,i\,t$$

If we consider a very small increment of time, dt, the capacitor will charge a correspondingly small increment of voltage, de_o. The output capacitor voltage, expressed in differential form, is then

$$de_o = \frac{1}{C}\,i\,dt$$

The ultimate output voltage will be equal to the summation of all the differential increments of charge prevailing during a specified time interval. Expressed in integral form:

$$e_o = \frac{1}{C}\int i\,dt$$

RC Integrator

The fact that the voltage across the capacitor is proportional to the charging current allows the RC circuit to be used as an integrator. Provision must be made, however, to supply a charging current that will be *proportional* to or *vary linearly* with the input voltage representing the information to be integrated. A direct proportionality exists between the current and voltage drop produced by the current flowing in a pure resistance; that is, $I = \frac{E}{R}$. Hence, by adding a resistor in series with the capacitor, a charging current *approximately* proportional to e_i will flow into C, provided the time constant (the product of RC) is made very large. *A Fundamental Circuit of an RC Integrator and Its Output Waveforms* are

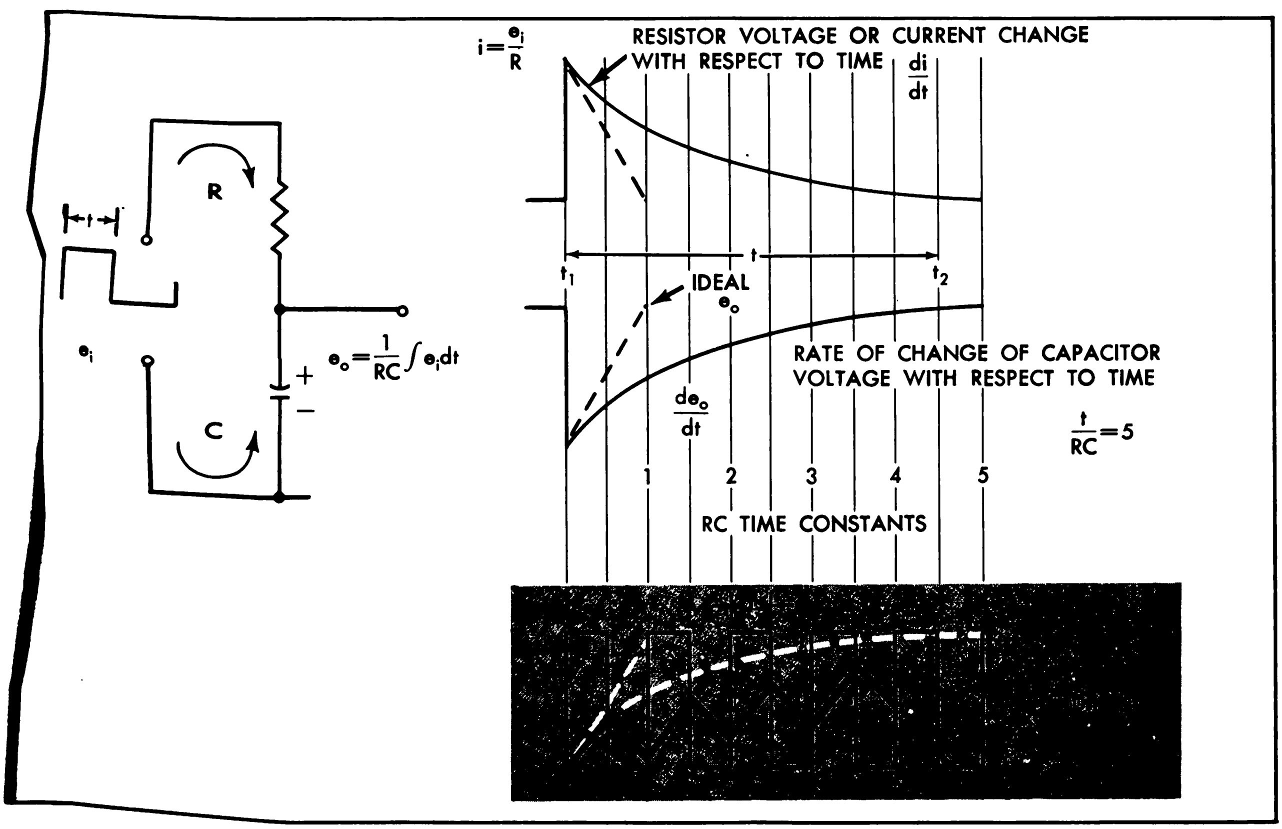
illustrated in the figure; here the capacitor voltage represents the integrator ouput.

As shown in the illustration, a rectangular voltage e_i is suddenly impressed upon the circuit at time t_1. At the instant this voltage is applied, the capacitor, which initially has no charge on it, acts like a short circuit, and the only opposition to the flow of current is the resistance which limits the current to its maximum value, $\frac{e_i}{R}$. At this instant, the entire applied voltage e_i is dropped across the resistor and the voltage drop (charge) across the capacitor is zero. As time progresses, current flows in the direction indicated by the arrows in the RC circuit, exponentially charging the capacitor toward the applied voltage, to the polarity shown. As the capacitor continues to charge, it offers an increasing opposition to the charging current; thus, the capacitor continues to charge at a decreasing rate, and the charging current becomes less proportional to the input signal.

After an interval of time (t_1 to t_2 — 5 RC time constants), the capacitor becomes fully charged, and the current flow stops completely. At the instant t_2, the current through and the voltage drop across the resistor have decayed to zero, and the charge on the capacitor has become equal to the applied charging potential.

Linearity of Integrated Output

The resulting deviation of the charging current or output voltage from strict proportionality or linearity is shown by the curves in this example. The ideal output e_o for a constant input signal would be increasing uniformly with time as indicated by the broken lines; note that this ideal



Fundamental Circuit of an RC Integrator and its Output Waveforms

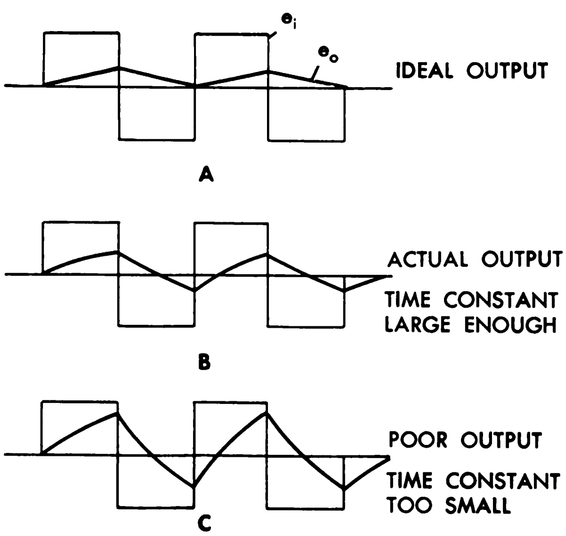
(linear) output is only approached during the beginning of the charge path before the capacitor has become appreciably charged.

One remedy to this deviation from linearity in an RC integrator is to use a circuit with a time constant that is large relative to the charging period of the applied voltage. That is, by making the ratio of $\frac{t}{RC}$ equal to 0.1 or less, we use only a very small percentage of the total charge curve. The result is a more accurate integration of the input signal as shown by the *inset* in the illustration of the fundamental *RC* integrator.

If the charging current i were strictly proportional to the input voltage e_i, the equation for the output voltage could be expressed as:

$$e_o = \frac{1}{RC}\int e_i dt = e_i \frac{t}{RC}$$

And this ideal output would be a perfect triangular wave as shown at A in the illustration *Comparing Ideal and Actual Outputs of an RC Integrator.* Note that, although a long time constant produces more accurate results, it also provides a much lower output



Comparing Ideal and Actual Outputs of an Ideal Integrator

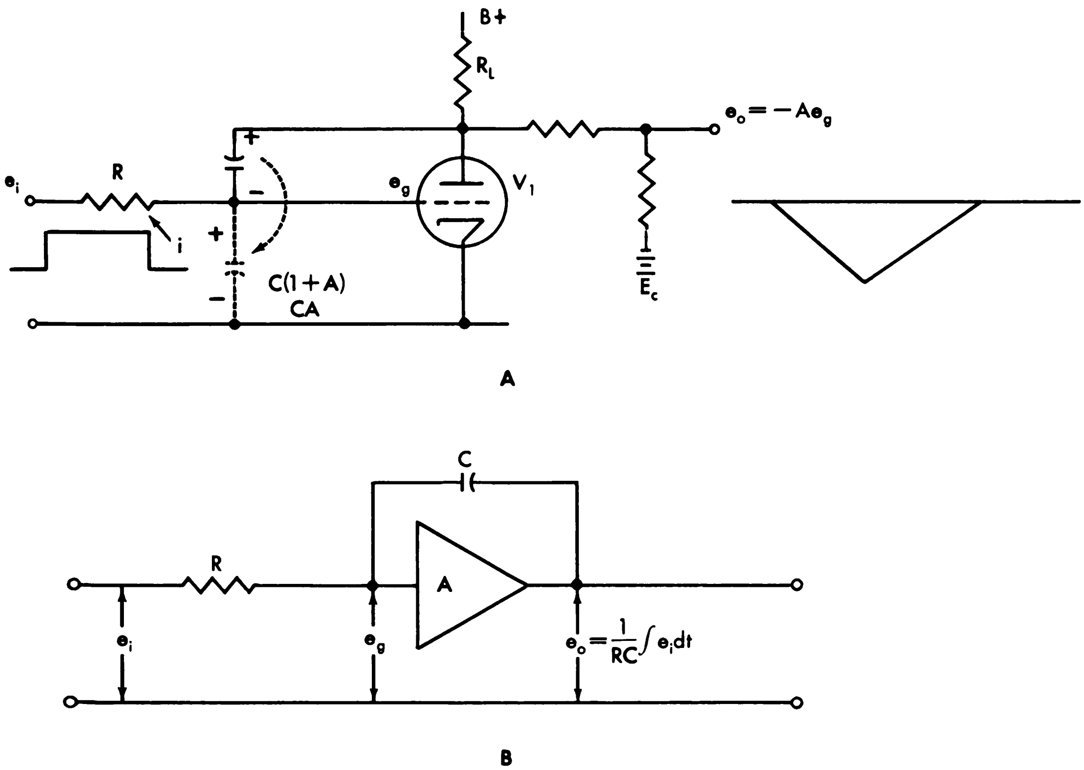
for the same input signal. However, good integration can be accomplished in a loaded RC network only at the expense of using large values of R and C to produce a long time constant, because the effect of loading is to decrease the time constant and, in turn, the accuracy of integration. The simple RC network has the obvious advantages of being simple, inexpensive, and free from drift because it contains no vacuum tubes.

Isolating the Load From the Integrating Network. The disadvantages resulting from loading RC integrating networks may be overcome to a great extent by using buffer amplifiers to isolate the load from the integrating capacitor. As in the case of summing and multiplier networks, the buffer amplifier may be of the cathode follower type or any other amplifier with very high input impedance and high voltage gain. A few of the special circuits utilizing special tubes drawing very little grid current will be discussed in greater detail in the section on amplifier components of computers.

Integrating Amplifier with Negative Feedback

The accuracy of integration may be improved with the use of a high-gain negative feedback (operational amplifier). The basic schematic A and the conventional symbolic representation B of an integrating amplifier with negative feedback are illustrated. Note that the circuits are similar to those of the sine-reversing or summation (operational) amplifier previously discussed except that the feedback resistor R_f has been replaced by the feedback capacitor C.

You may recall, from your previous study of amplifiers, that any capacitance present between the plate and grid of a vacuum tube is reflected back into the grid circuit and added to the interelectrode capacitance between the grid and cathode. This effect was first described by Miller and has since been known as the Miller effect; hence, the circuit is commonly referred to as a Miller integrator. The Miller effect proposes that the feedback capacitance reflected back to the



Integrating Amplifier with Negative Feedback

grid acts like a virtual capacitance (between the grid and cathode) which is equal to $C\ (1 + A)$. Since the voltage gain A is very large, we may neglect unity, and the virtual integrating capacitance becomes $C A$.

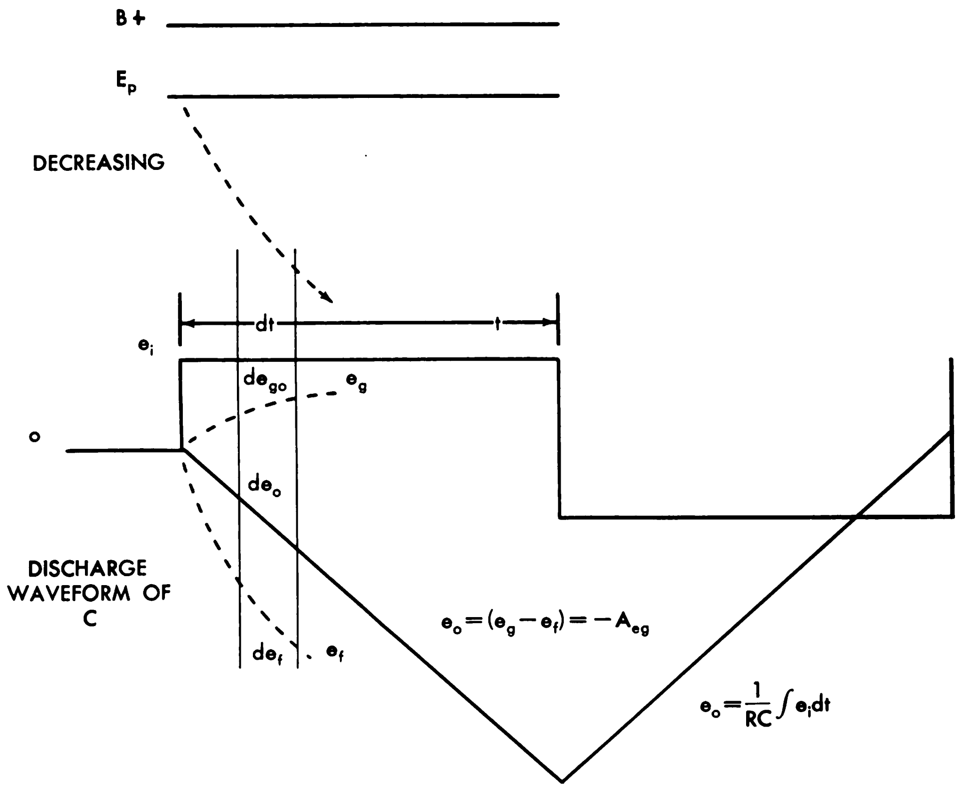
OPERATION. The operation of this circuit can be explained by assuming a rectangular input with a constant amplitude as shown. At the start, assume the initial conditions to be zero; that is:

$$e_1 = e_g = e_o = 0$$

Also assume that the feedback capacitor is charged as indicated to the static-state voltage between plate and grid.

When the positive step input pulse to be integrated is applied, the capacitance $C A$ suddenly begins to charge through R toward e_i with the polarity indicated in the illustration. This increment of charge e_g is coupled to the input of the DC amplifier where it is amplified, inverted, and fed back through the feedback capacitor so as to oppose the charging of $C A$. In other words, e_g tends to rise in the positive direction which makes the amplifier tube conduct more heavily. Hence, the plate to cathode voltage of V_1 decreases, causing C to discharge and oppose the charging of $C A$.

OUTPUT WAVEFORMS. When examining the waveforms of the grid and feedback voltages (next illustration), you can readily see that the curvature of the grid voltage is also opposed by the curvature of the feedback voltage, thus allowing the output voltage to be very linear. Note that the output voltage is negative and equal to the grid voltage minus the feedback voltage and that it changes A times as fast or steeper than e_g. The magnitude of e_g has been exaggerated here for simplicity of illustration. Actually



Integrating Amplifier Waveforms

for the constant input voltage e_i, the charging current of $C A$ is limited by the feedback to a particular linear value which tends to keep e_g practically at zero.

OPERATIONAL EQUATION FOR THE MILLER INTEGRATOR. At any instant, the charging current for this virtual integrating capacitor (CA) must pass through the input resistor. This current then will be equal to the voltage across the input resistor divided by the resistance; that is,

$$i=\frac{e_i}{R}$$

The voltage across this capacitance at any instant will be equal to the quantity of charge divided by the capacitance; that is,

$$e_g=\frac{Q}{CA}$$

or, since $Q = it$,

$$e_g=\frac{it}{CA}$$

where i is the charging current.

Substituting the quantity $\frac{e_i}{R}$ for i in the above equation,

$$e_g=\frac{e_i}{R}\,\frac{t}{CA}=e_i\,\frac{t}{(RC)A}$$

Hence, when a negative feed back amplifier with high gain is used, the time constant of integration is effectively increased by the factor, A. Thus, the Miller integrator is able to integrate a step input pulse for a period A times as long as the fundamental RC network without the amplifier with the same degree of accuracy.

Since $e_o = -A\,e_g$, multiplying the above equation through by $-A$ gives the equation

for the output voltage; that is,

$$e_o = -Ae_i \frac{t}{RCA}$$

$$e_o = -e_i \frac{t}{RC}$$

Considering a very small increment of time dt, the output will change by a correspondingly small increment of voltage de_o, that is,

$$de_o = -\frac{1}{RC} e_i \, dt$$

The ultimate output voltage will be equal to the summation of all the differential increments prevailing during a specified time interval. Expressed in integral form,

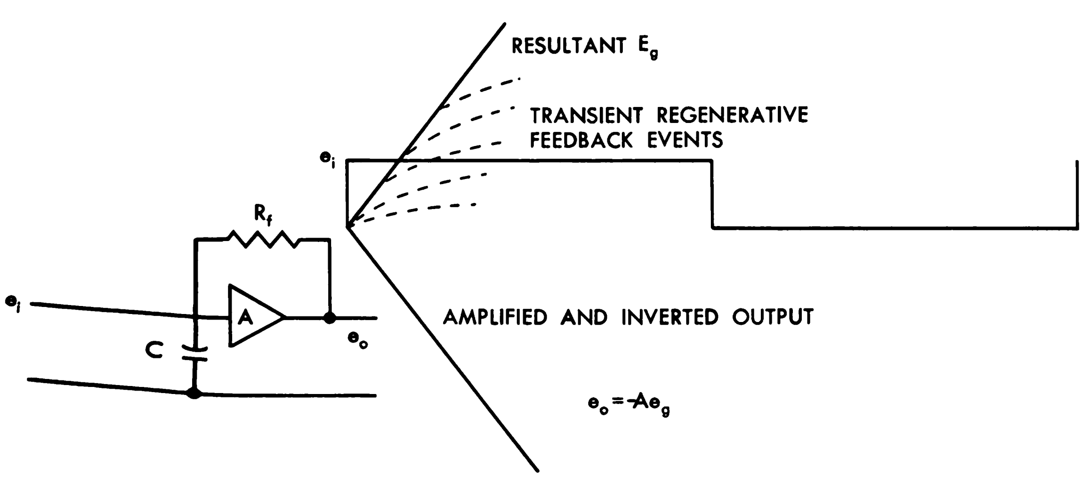
$$e_o = -\frac{1}{RC}\int e_i \, dt$$

Integrating Amplifier With Positive Feedback

The accuracy of integration may also be improved by providing positive feedback in phase with the original charging potential. If we provide the proper positive feedback to the basic RC integrator, the input voltage to be integrated is boosted by an amount at the proper time to overcome the nonlinearity of the charging current. An integrating amplifier with positive feedback and its accompanying waveforms are illustrated. You will find that the operation of this circuit is similar in principle to the bootstrap sweep generator used to provide a linear sweep voltage for fire control radar.

In this circuit the familiar exponential form of voltage appearing across the integrating capacitor is amplified by the DC amplifier. The output voltage from the amplifier is then fed back through the feedback resistor and added to the original integrator input so as to neutralize the nonlinear charging rate of the capacitor. That is, a sequence of transient (amplifying and feedback) events is repeated over and over, increasing the available charging potential to a continuously higher value; as a result the capacitor continues charging at a linear rate.

The waveform indicated as e_g in the illustration shows the resultant linear rate of accumulation of charge on the integrating capacitor resulting from the accumulation of the transient regenerative feedback events. If the amplifier gain and feedback are properly chosen, the output, as well as the voltage across the capacitor, will be proportional to the integral of the input voltage to be integrated. Note that the illustration shows the output of the regenerative feedback integrator to be an amplified and inverted version of the voltage appearing across the integrating capacitor.



Integrating Amplifier with Positive Feedback

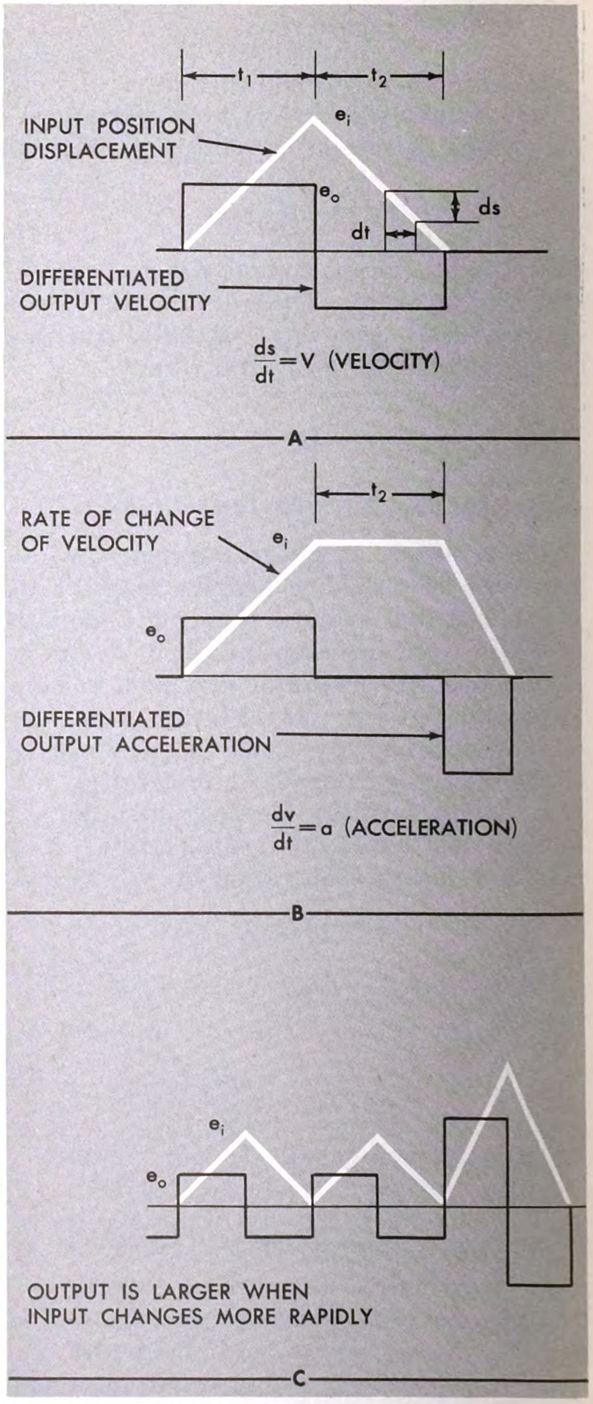
SECTION D, DIFFERENTIATOR COMPONENTS OF COMPUTERS

A differentiator is part of the rate system of a computer—it, like the integrator, performs a mathematical operation on an input signal. The input signal may be some type of information such as position error, angular deviation, altitude, velocity, or control surface position. The mathematical operation that determines the rate at which an input signal is changing is called differentiation. A differentiator is a device which performs this mathematical operation called differentiation and produces an output proportional to the *rate of change* of magnitude of the input signal.

REACTION OF AN IDEAL DIFFERENTIATOR

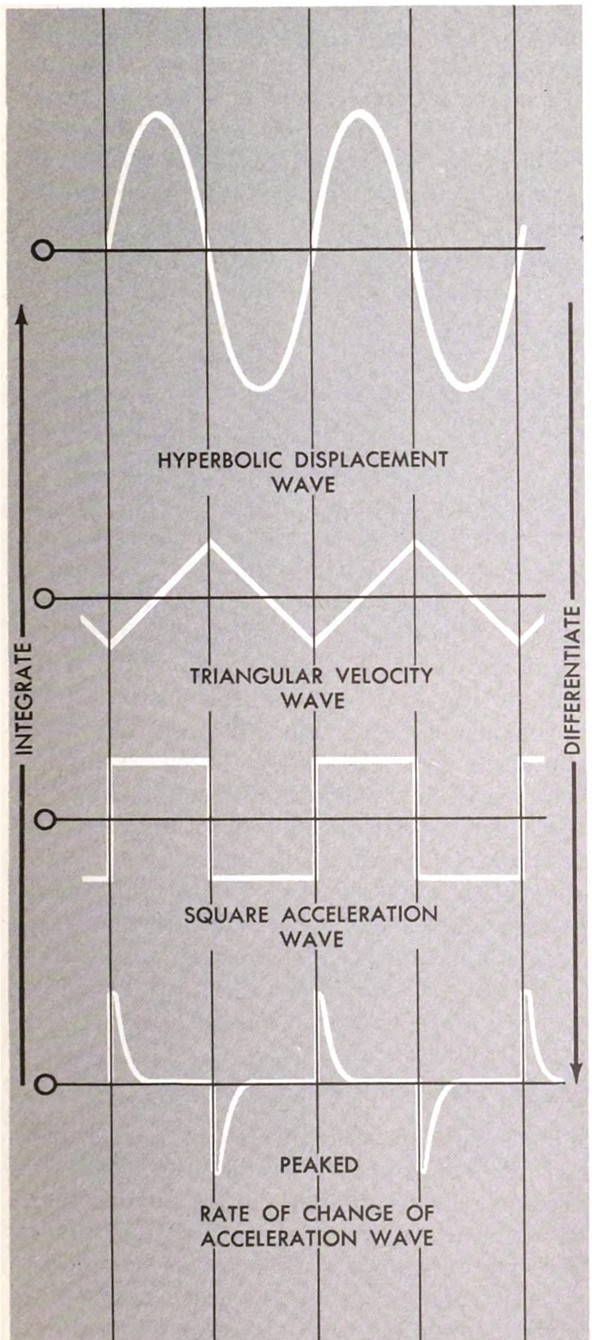
If you carefully examine the waveforms illustrating the *Reaction of an Ideal Differentiator to Various Input Signals,* you will see how the magnitude and sense of the differentiator output depends on the sense and rate of change of the input signal. For example, at **A** in the illustration, the input voltage (representing the rate of change of position displacement) increases at a linear rate during the time interval t_1. The derivative of this displacement distance with respect to time $\left(\frac{ds}{dt}\right)$, or the output of the differentiator, is a step voltage with a constant magnitude representing a constant velocity during this time interval.

During the next time interval t_2, the input voltage representing position displacement is shown to be decreasing toward zero at the same linear rate of change. At the beginning of this second time interval, the differentiated output voltage suddenly reverses sense with the same magnitude and for the same period of time, thus representing a constant velocity which is now in the opposite direction. This shows that the derivative (differentiated output) of a triangular input wave *representing displacement* is a square



Reaction of an Ideal Differentiator to Various Input Signals

wave *representing velocity.* If the input voltage represented the rate of *change of velocity,* the differentiated output would represent the *magnitude and sense* of acceleration during the periods t_1 and t_2. You will recall that the rate of change of velocity with respect to time is equal to acceleration —that is $\frac{dv}{dt} = a$. The waveforms at **B** illustrate how the output of the differentiator should remain at zero as long as the input signal remains constant. That is, even though the input signal is large in magnitude during the interval of time designated as t_2 (representing a high but constant velocity), the differentiated output (representing acceleration) should remain at zero.



Comparing Differentiation with Integration

The waveforms at **C** in the illustration show that, if an input signal (representing either the rate of change of displacement or velocity) is increasing or decreasing slowly, the output of the differentiator (representing the sense and magnitude of either velocity or acceleration) is small. However, if the magnitude of the input signal changes more rapidly, the differentiated output is proportionally larger.

All waveforms in the illustration show that the differentiated output reverses sense when the input changes from an increasing to a decreasing signal.

Comparing Differentiation with Integration

If you will now reexamine the waveforms illustrating the reaction of an ideal integrator on page 4-32 and compare them with the waveforms (just discussed) illustrating the *Reaction of an Ideal Differentiator,* you will find that the process of differentiation is the reverse of integration. This principle is summarized and reemphasized in the next illustration, *Comparing Differentiation with Integration.* Note that the process of changing the square wave (representing acceleration) into a triangular wave (representing velocity) is called integration. When this triangular wave is integrated, the result is a *hyperbolic* waveform representing *position displacement* or distance off-course. Notice that this curve is *not* a sine wave but is a series of hyperbolic curves which form a periodic waveshape. This hyperbolic wave-

form may represent the undesirable deviation in *yaw* or oscillation about a true heading due to *overcontrol* of an order signal which commands system action.

If the hyperbolic waveform is differentiated, the result is a triangular waveform as shown. If the triangular waveform is then differentiated, the result is the square wave. The peaked wave at the bottom of the illustration represents the rate at which acceleration is changing and is the result of differentiating the square wave.

You will recall that in an RC *integrating* circuit (one that produces an output that is the integral of the input) the time constant had to be very long and the output was taken across the capacitor. You will find that in an RC *differentiating* circuit the time constant must be very short and the output must be taken across the resistor.

Differentiating or integrating a sine wave by an RC circuit produces another sine wave having a different amplitude and phase, but with the same sinusoidal waveshape. The result of differentiating a sine wave representing displacement will be illustrated next and discussed on page 4-46 while investigating attitude differential control.

PURPOSE OF DIFFERENTIAL CONTROL

Differentiators are required in many interceptor control computers in order to improve *rate stability*. When an order signal which commands system action is applied, it is very important that it be applied at the proper time, for the proper length of time, and with the proper magnitude. Differentiators are necessary because of the fact that a definite amount of time is required to perform any type of control operation such as proper attitude, guidance, and fire control.

For example, the correction for deviation off-course cannot occur at the instant that the position displacement is sensed, for this would result in no deviation whatsoever. Such a condition would be ideal, but it is impossible to attain since the system must detect some error in order to operate. However, since the ideal condition described would have zero lag time and would result in no deviation whatsoever, efforts are made to approach the ideal condition and reduce the lag time to a minimum.

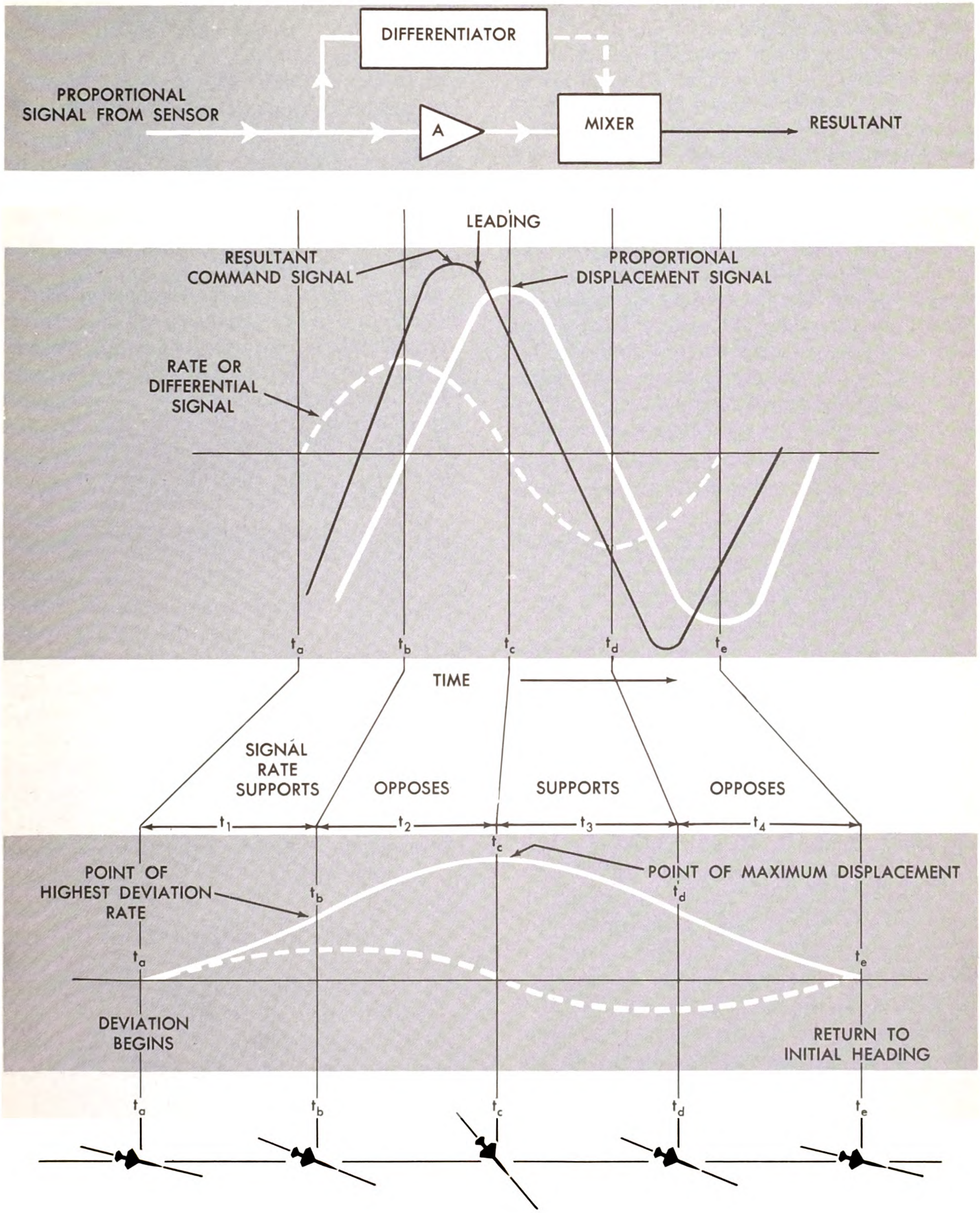
The airframe and control surfaces of interceptor aircraft and missiles are designed to correct attitude rapidly. When an order signal which commands system action is applied, the control surfaces are moved by powerful, fast-acting actuators using an amplified error signal to produce fast response. However, the reduction of lag by such measures approaches a limit beyond which it cannot go.

Oscillation Due to Overcontrol

At this point, you may be thinking, "Why not increase the amplification of the error displacement signal (order signal which commands system action) so that even a very small sensed deviation will produce a signal to correct attitude almost instantaneously?" This ideal is all right up to a certain point. However, a signal which is too large returns the craft beyond the desired point in the opposite direction. The resulting opposite error signal causes the craft to deviate back again in the original direction. The end result is the undesirable oscillation about the desired true heading—as was shown by the hyperbolic waveform in the last illustration. Furthermore, without a differentiator, a rapid deviation would normally produce a fairly large displacement due to the lag which occurs before correction takes place.

Differentiator Action in a Rate Computer

The addition of a differentiator in a control system computer enables the system to predict a large possible future displacement from the existence of a rapid rate of deviation and commands the proper corrective action. The output of the differentiator is usually combined with the proportional displacement signal to produce a resultant signal which leads the original proportional


DIFFERENTIATOR
PROPORTIONAL SIGNAL FROM SENSOR
A
MIXER
RESULTANT
LEADING
RESULTANT COMMAND SIGNAL
PROPORTIONAL DISPLACEMENT SIGNAL
RATE OR DIFFERENTIAL SIGNAL
t_a t_b t_c t_d t_e
TIME
SIGNAL RATE SUPPORTS
OPPOSES
SUPPORTS
OPPOSES
t_1 t_2 t_3 t_4
POINT OF HIGHEST DEVIATION RATE
POINT OF MAXIMUM DISPLACEMENT
DEVIATION BEGINS
RETURN TO INITIAL HEADING

signal. That is, the resultant order signal that commands system action is advanced in time, resulting in an apparent reduction in the overall lag of the system. The end result is a reduction in overall displacement, a minimum of overcontrol, and damping of oscillation about true heading.

ATTITUDE DIFFERENTIAL CONTROL. A block diagram and its associated waveforms illustrating the combination of a proportional signal with its differentiated component—to produce a leading command signal—is shown on page 4-45. The proportional signal is assumed to be a sine wave as indicated by the solid white waveform which represents the magnitude of displacement or position off-course during the specified time intervals.

Note that part of the proportional signal is fed to the input of the differentiator which produces a signal representing the rate of change of displacement. This differentiated output (the rate signal indicated by the broken white lines) is then combined with the amplified portion of the proportional signal to produce the resultant leading command signal.

The waveforms show that, as the interceptor begins to deviate from the desired true heading, the rate signal supports the corrective action of the displacement signal. For example, at time t_b the proportional signal has its steepest slope, the off-course displacement is changing at its maximum rate, and the rate signal is at its maximum amplitude, all of which indicate a maximum rate of change. At this instant in time, when the displacement from true heading attempts to increase at a rapid rate, the differentiated signal is large and supports the proportional signal. This supporting rate signal is added vectorially to the proportional signal, producing a resultant leading command signal which reduces the rapid deviation.

After corrective action has been applied for the time interval designated as t_1, the interceptor deviates less rapidly, and we find the rate signal decreasing in magnitude. Hence, during the interval t_2 the rate signal opposes the displacement signal. Subsequently, as indicated by the proportional signal (at t_c), the off-course displacement has reached its maximum value. At this instant the output from the differentiator is zero because the interceptor is not changing heading and the proportional input signal to the differentiator has no rate of change. Remember there is a rate output only when deviation is changing.

An instant later, as the interceptor begins to return to its initial heading, the proportional displacement signal begins to decrease. Hence, the differentiated signal reverses sense and again supports the displacement signal during the interval t_3. Finally, the interceptor returns to its initial heading as shown at t_c, with the rate signal opposing the displacement signal during the interval t_4. If the rate and displacement signals were to support each other during the interval t_4 as the interceptor returns to the desired heading, the interceptor would return so fast that it would swing past the required position in the opposite direction. This, of course, would cause overcorrection with the end result of oscillations about the desired heading being enlarged rather than damped. However, the rate and displacement signals are mixed, as shown, with the proper rate and phase so as to actually support corrective action at all times and keep oscillation to a minimum.

ELECTRONIC DIFFERENTIATOR

RC combinations which are used in fundamental RC integrators may also serve as differentiators—if the RC time constant is made much smaller and the output is taken across the resistor. You will recall that the output voltage developed across the resistor in a series RC circuit is proportional to the charging current. Charging current can be made approximately proportional to the rate of change of applied voltage. Hence, the *magnitude* of voltage across the resistor may

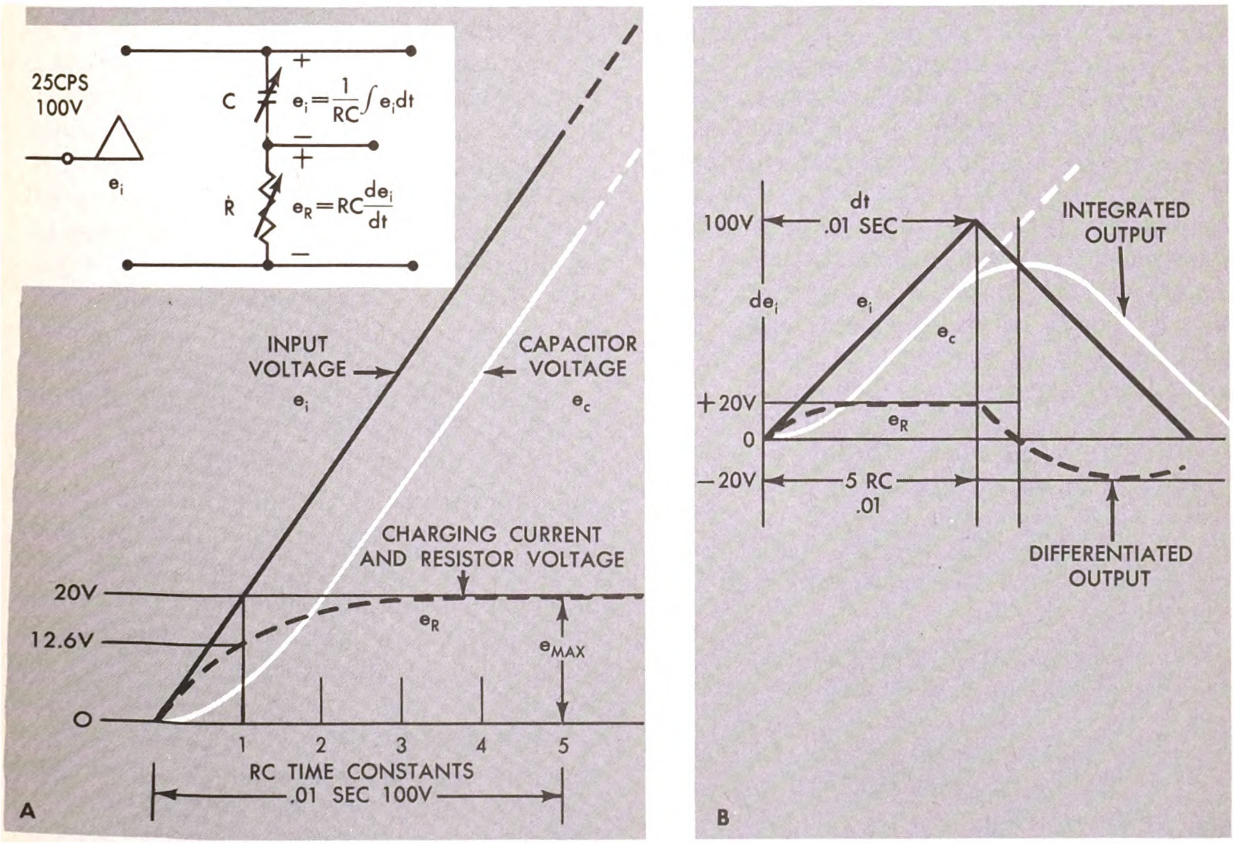
be used to represent the *rate of change* of input voltage.

The Fundamental RC Differentiator

For ease of understanding the operation of a fundamental RC differentiator, let us assume a triangular input voltage which has a constant rate of change during alternate half-cycles. Let us first examine the effects of the first half-cycle when a triangular wave is applied to a fundamental RC differentiator. The waveforms at **A** in the diagram illustrating *Response of an RC Circuit to a Triangular Input Voltage* show a small portion of the first half-cycle enlarged several times for easier examination.

THE FIRST HALF-CYCLE. Notice that the input voltage starts at zero and keeps on rising at a constant rate. As this input voltage increases, the capacitor must continually charge toward a new higher value.

Initially, when the voltage e_i is applied, the charging current increases rapidly with an increase in applied voltage because the capacitor is initially uncharged. However, this current stacks electrons on the capacitor plate which charges it to the indicated polarity so as to oppose the applied voltage and the charging current. However, since the input voltage is continually increasing, the opposing capacitor voltage is more than overcome by the increase in applied voltage. Hence, the current actually continues to increase—but less rapidly than initially. As current continues to increase, the capacitor becomes charged to a higher voltage and the rate of current increase falls off more and more. Finally, at the end of five RC time constants, the charging current is large



Response of an RC Circuit to a Triangular Input Voltage

enough to raise the capacitor voltage at the same rate at which the input voltage is increasing. Note that, at this point, the current stops increasing and becomes constant and that the slope of the capacitor voltage becomes the same as the input voltage slope. Also at this time, the constant current through the resistor causes a constant voltage to be dropped across the resistor.

Since the current increases from zero to a constant value during the first five time constants, its rate and the rate of change of resistor voltage is exponential. Hence, e_R at the end of one time constant is equal to 63.2% of the voltage applied at that instant. After five time constants, e_R levels off to a value equal to the voltage applied at the end of the first time constant—that is, to the value designated as e_{max}.

THE SECOND HALF-CYCLE. The waveforms at A and B in the illustration show that, after five time constants, the capacitor would continue to charge at the same linear rate until the input voltage ceases to rise. However, as the input voltage abruptly starts decreasing during the second half-cycle, the charging rate of the capacitor decreases. Note that the charging rate decreases until the input voltage drops and becomes equal to the voltage to which the capacitor has charged. At this point, the current is zero. As the input voltage continues to decrease, it falls below the capacitor voltage, and the capacitor starts discharging. This reverses the current and produces a negative voltage across the resistor. Again, the changeover cannot occur any faster than the capacitor can discharge, so the current change and consequently the resistor voltage change are exponential.

THE MAGNITUDE OF DIFFERENTIATED OUTPUT. The maximum value of resistor voltage (e_{max}) is equal to the product of the rate of change of input voltage and the RC time constant. It may be expressed in differential form by the equation,

$$e_{max} = RC \frac{de_i}{dt}$$

For example, assume that the triangular input wave shown in the illustration at B has a frequency of 50 cps and a peak input voltage of 100 volts. Inspection shows that the change of input voltage (de_i) during the first half-cycle is 100 volts. Since the frequency is 50 cps, the duration of the complete cycle is $\frac{1}{50}$ or .02 second. Therefore the duration of the half-cycle designated as dt is equal to .01 second. If $R = 100K$ and $C = .02$ ufd, the time constant

$$RC = 10^5 \times .02 \times 10^{-6} = .002 \text{ sec}$$

Substituting these determined values in the equation,

$$e_{max} = RC \frac{de_i}{dt} = .002 \left(\frac{100}{.01}\right) = 20 \text{ volts}$$

EFFECT OF VARYING THE TIME CONSTANT. In the example just discussed the five RC time constants were made equal to the duration of one half-cycle. That is, the resistor voltage just barely reached its maximum constant value of 20 volts within the period dt in which the input voltage was increasing. However, upon close reexamination of the output equation and the waveforms in the illustration, it can be seen that, if the time constant is shortened, the condenser will reach its maximum linear charging rate in less time. Hence, the resistor voltage would reach its maximum in a much shorter percentage of the total period dt in which the input voltage is increasing. It can be seen that decreasing the time constant allows the resistor voltage to more nearly approach a square wave—the ideal differentiated output with a triangular input which was discussed earlier in this section. However, a fairly accurate differentiated output is obtained if the time constant does not exceed one-fifth the duration of one half-cycle— as in the case of the example discussed and illustrated. Note also, that with parameters used in the example, the integrated output (across the capacitor) very nearly approaches an ideal hyperbolic waveform.

MECHANICAL DIFFERENTIATOR

Differentiating any function with respect to time is simply speed measurement; hence, any tachometer or rate measuring device is a time differentiator. The automobile speedometer is an example of this. The input of a speedometer may be a rotating shaft with some sort of rate sensing device incorporated to detect and indicate speed.

The rotational velocity of the shaft can be detected by centrifugal force or by magnetism. Suppose a weight is fastened to a rotating member so that it can swing outward against spring tension. The amount of centrifugal force tending to pull the weight from the shaft against spring tension depends on the speed of rotation. Hence, the displacement of the weight, the extension of the spring, or the spring tension itself could be a measure of the rotational velocity.

A special generator could also be used to indicate rotational velocity. Since the rate at which magnetic lines of force are cut determines the output voltage, a voltmeter could be calibrated in rpm, miles per hour, or other suitable units to indicate rotational velocity.

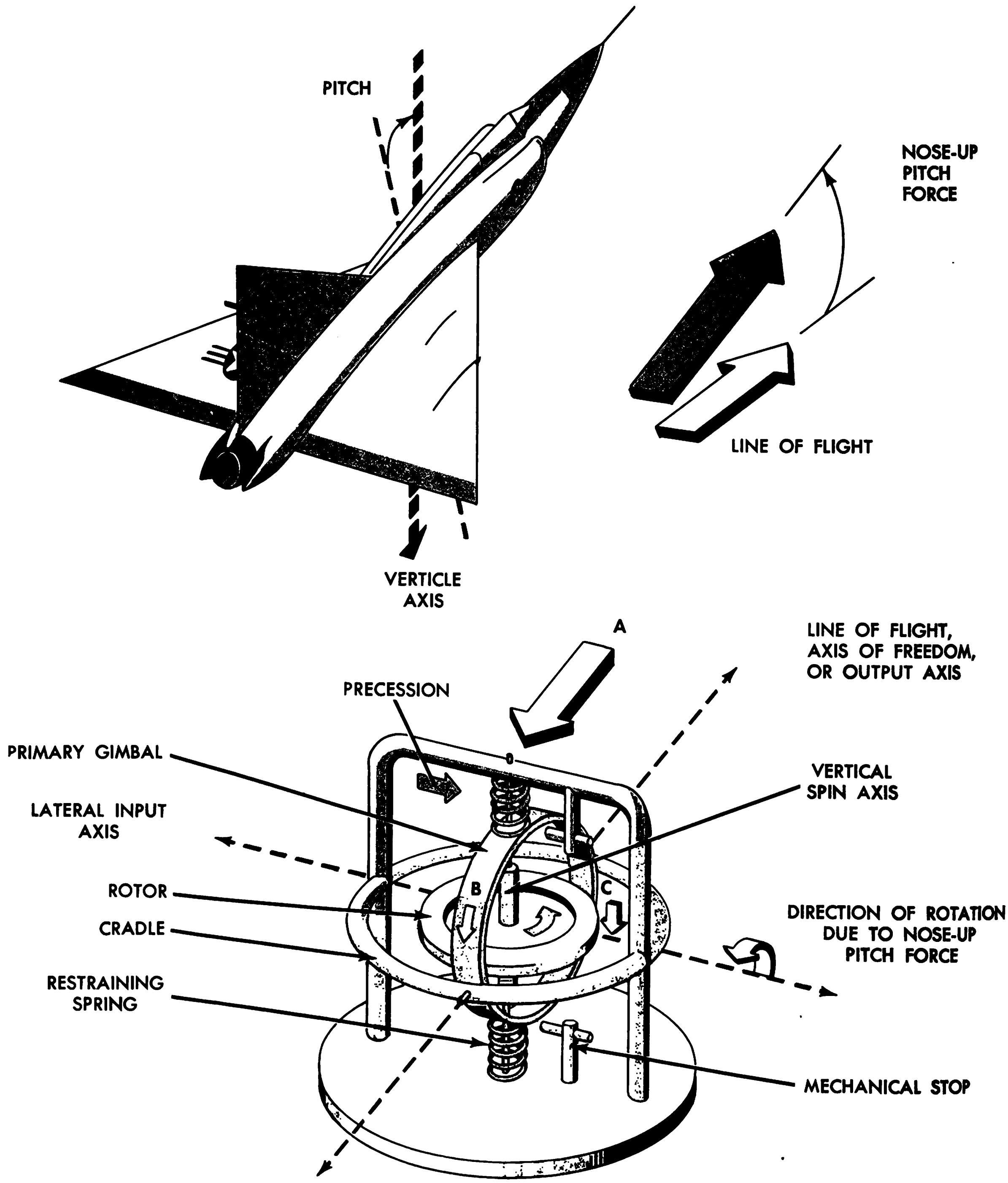
Differentiating or Rate Gyro

A common method of differentiating an angular displacement with respect to time and producing a rate signal proportional to the rate of change of deviation is by means of a rate gyro. A rate gyro is a special type gyro which is similar to a free gyro in that it has a high gyroscopic momentum while its rotor is spinning. It differs, however, in that it is free to rotate about one axis only and that it is sensitive only to angular rates of movement about an axis perpendicular to the spin axis. In other words, an angular displacement of its cradle (above a certain rate about its input axis) will cause its spin axis to change position in relation to space. Such a change will result in precession. Precessional motion about the axis of freedom is restrained by a linear spring. Hence, a force is exerted on the spring which is proportional to the angular rate of movement about the input axis. Thus, by measuring the precessional force exerted on the spring, an indication of the rate of deviation may be obtained directly. This precessional force may be detected with a microsyn pickoff generator mounted on the output shaft of the rate gyro.

DIFFERENTIAL CONTROL SYSTEM EMPLOYING A RATE GYRO. A differential control system employing a rate gyro usually requires a separate rate gyro for every plane of motion in which a rate signal is desired. Normally, a yaw rate gyro is mounted with its spin axis parallel to the longitudinal axis or line of flight. A roll rate gyro is mounted with its spin axis parallel to the lateral axis, and a pitch rate gyro is mounted with its spin axis parallel to the vertical axis.

The Pitch Rate Gyro illustrated may be used to demonstrate the action of a rate gyro when used in an interceptor differential control system. Note that the gyro is mounted so that its primary gimbal has only one degree of rotational freedom about the longitudinal axis which is parallel to the line of flight. If the interceptor noses upward, the gyro housing or cradle (since it is secured to the air frame) is forced to revolve about the input lateral axis. This angular force may be represented as a pitch force applied to the cradle at either *A* or *B*. The downward force at *B* is transmitted 90° in the direction of rotor spin and causes a downward precessional force at *C*. This precessional force at *C* results in a subsequent precession which tends to force the top of the gyro gimbal to your right against the restraint of the spring. The gyro gimbal will precess about the longitudinal axis freedom until the torque imposed by the spring is sufficient to produce a rate of precession about the input lateral axis equal to that which it is actually being forced to move. Since the spring is linear, the precession about the lateral axis is proportional to rate of movement about the input pitch axis. As soon as the angular

rate applied to the cradle about the input axis ceases, restraining springs restore the gyro to its neutral position. The mechanical stops shown in the illustration are provided to limit the precession of the gimbal to a few degrees in either direction.



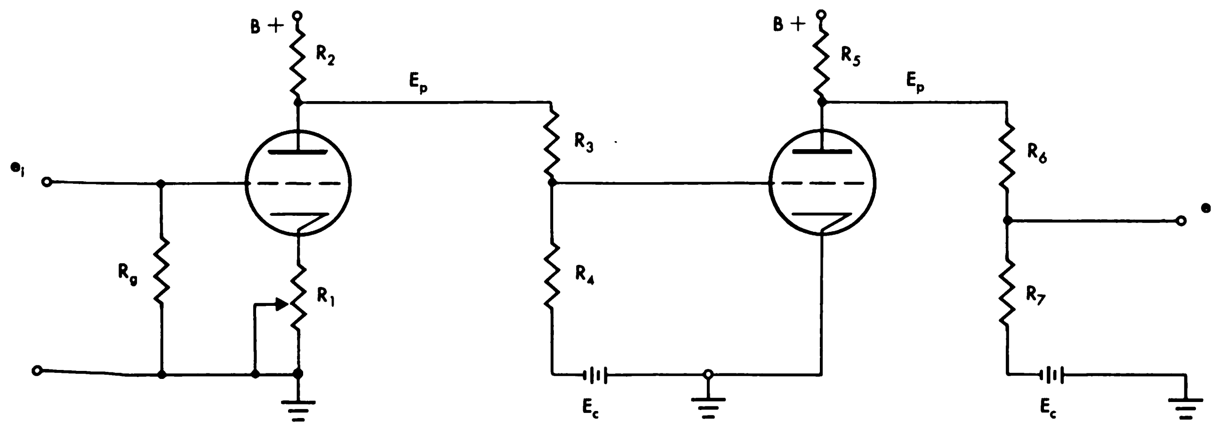
Pitch Rate Gyro

SECTION E, AMPLIFIER COMPONENTS OF COMPUTERS

One of the most important items used in analog computers is the high-gain direct-current amplifier. The sign changing and summing operational amplifiers discussed in previous sections may be of this type. Direct-current amplifiers are used to amplify very low frequency or DC voltage representing computer variables. The DC voltage that is to be applied must often be applied directly to the grid of the amplifier tube. In such cases, only direct coupling can be used in the amplifier input circuit.

DIRECT-COUPLED DC COMPUTER AMPLIFIER

One version of a two-stage direct-coupled DC amplifier is illustrated.



Basic DC Amplifier

In this basic DC amplifier circuit, a DC voltage may be applied directly to the grid of the input stage. The amplified output—the plate voltage—of this first stage is directly coupled to the grid of the second stage.

The special nature of DC amplifiers requires that bias and plate voltage on each succeeding stage be greater than the bias and plate voltage of the preceding stage. This makes cascading of DC amplifiers rather difficult and makes necessary the use of at least two regulated power supplies.

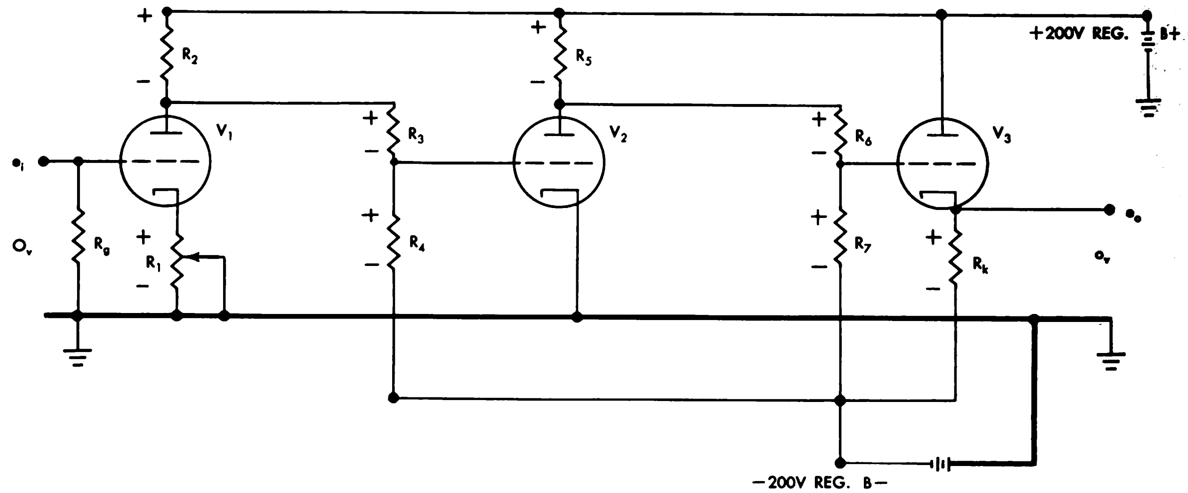
Input and Output Reference Levels

In order to interconnect various computing elements and to make it possible to apply the proper feedback from the amplifier output terminals, practically all DC amplifiers used in DC analog computers must have an output voltage reference level. This reference level is usually zero with respect to ground.

The desired output reference level may be obtained by means of a resistor summing network similar to that shown in the illustration of the *Basic DC Amplifier*. Note that this network, consisting of R_3 and R_4 or R_6 and R_7, permits a bias voltage E_c to be added to the voltage appearing at the output (plate) of either tube—thus allowing any desired DC reference level to be obtained at either the output or input of the second stage.

To further describe the characteristics of the basic DC computer amplifier, reference

will be made to the next illustration *A Basic Computer Amplifier with Cathode Follower Output*. You will note that the coupling between each plate and the following grid is again by the same familiar resistor chain. However, the lower end of these summing resistors (R_4 and R_5) is now connected to a common regulated negative bias supply rather than to the individual bias voltage (E_c) as previously illustrated. Upon further investigation you will find that the previous circuit and the present one are electronically identical except for the cathode follower output stage. The values of all resistors in the above circuit are adjusted so that, in the absence of an input signal, each grid is slightly negative with respect to its corresponding cathode, thus providing the required bias for normal operation. In the absence of an input signal there is, of course, no voltage dropped across R_g. Therefore, the input grid is at ground potential, and the first stage bias is provided by the unbypassed cathode resistor. For the second stage the bias is the algebraic sum of the —200 volt supply and the voltage appearing across R_4. The bias for the cathode follower output stage is equal to the algebraic sum of the voltage drops across R_7 and R_k. It must be remembered that the output terminal of the cathode follower must be at zero potential with respect to ground in the absence of an input signal to the amplifier.



A Basic Computer Amplifier with Cathode Follower Output

Manual Zero Adjustment or Balancing

The output level of a DC amplifier is subject to drift due to changes in plate or bias supply voltages, changes in resistance values, and vacuum tube characteristics caused by temperature changes and aging. This drift results in an unbalance between input and output terminals and leads to errors in the desired computations of computer amplifiers. Before each computer run, it may be necessary to adjust one or more of the resistors affecting the voltage levels so as to obtain the proper output voltage with respect to the input. This procedure is known as ***zero adjustment*** or *balancing*. Most amplifiers are fitted with some specific device such as an adjustable bias control by which this zero adjustment can be accomplished before each computation begins. For example, in the amplifier just illustrated, zero adjustment is accomplished by short-circuiting the input

terminals and adjusting R_1 until the output voltage at the cathode of V_3 is equal to zero.

There are, of course, numerous other variations of DC computer amplifiers, but they are all based on the circuits previously shown. To obtain a desired reversal of sign between input and output, an odd number of stages are utilized. Since a single stage usually provides insufficient gain for most practical purposes, the three-stage version seems to be the most popular arrangement. Computer amplifiers may depend on manual reset and zero adjustment, or they may be of the drift corrected type. Some of the more complex computer amplifiers (to be discussed later in this section) include automatic balancing and zero adjustment circuits.

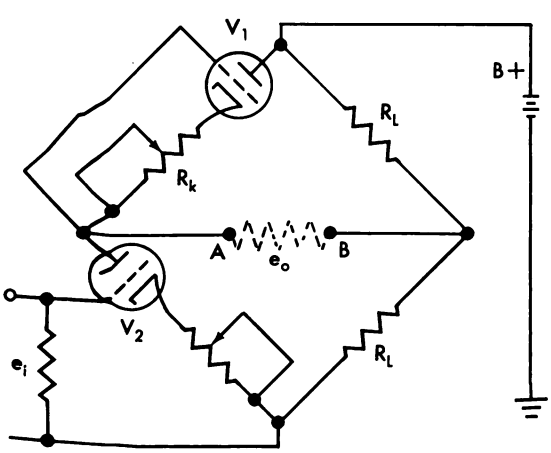
Drift Corrected Computing Amplifiers

A drift corrected amplifier is a development of a multistage amplifier where additional equipment is utilized to reduce drift. Drift is one of the most objectionable characteristics of all vacuum tubes.

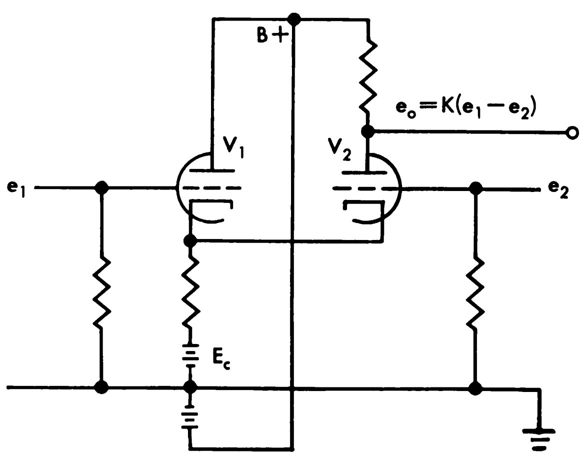
BRIDGE BALANCING PRINCIPLE. A common method of compensating for drift due to power supply changes in DC amplifiers uses the balancing bridge principle. As illustrated in the *Series Compensated Type Amplifier,* the output voltage is taken between terminals *A* and *B*. By adjustment of either of the cathode resistors, this output voltage may be adjusted to any desired reference level by increasing or decreasing the degree of unbalance. The output may be adjusted to zero—for zero input—by balancing the bridge. The output voltage reference level is particularly insensitive to changes in plate supply voltage. In fact, the plate supply voltage may be switched on and off without changing the output reference level. The tubes V_1 and V_2 should be as closely identical as possible for best cathode emission compensation.

DIFFERENTIAL COMPUTER AMPLIFIER. Since filament voltage fluctuations seem to have the same effect upon an amplifier output as an input signal, good quality DC amplifiers usually include some means of compensating for drift due to changes in filament voltage.

Two specially matched double triodes, V_1 and V_2, having a low negative grid current rating are utilized in the DC amplifier circuit illustrated. If the triodes in the circuit have a high amplification factor and a high value of cathode resistance is used, the circuit will produce an output voltage that is approximately proportional to the difference of the two input voltage changes. For this reason, the circuit is referred to as a differential amplifier.



Series Compensated Type Amplifier

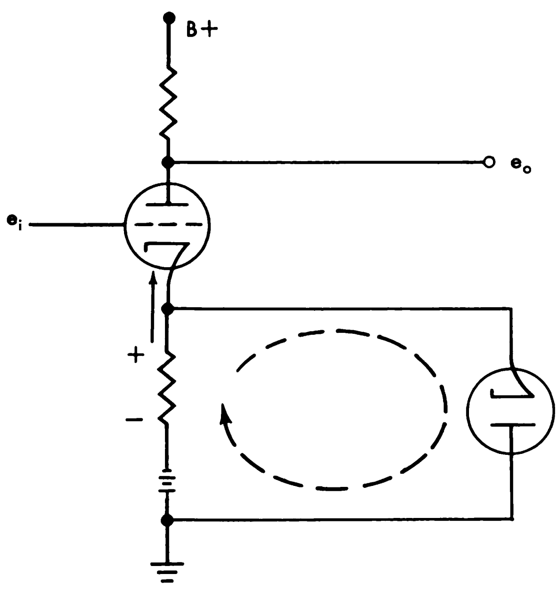


DC Differential Computer Amplifier

Variations of differential amplifiers are very useful in many circuit applications. They seem to be the most commonly used circuits for drift compensation due to variations in filament voltage and changes in cathode emission. They are also used to compensate for the effects of changes in plate and bias supply voltages.

Since the change in output voltage of a differential amplifier is approximately proportional to the difference between its two input voltage changes, changes of input voltages due to variations in power supply voltage (which are common to both input circuits) have a tendency to cancel out. Changes in cathode emission due to filament voltage fluctuations are virtually equivalent to changes common to both input terminals. They are, therefore, similarly cancelled out.

DIODE COMPENSATION. Filament voltage variations in a DC computer amplifier may be partially compensated for by placing a diode—having the same filament voltage as the amplifier tube—in the cathode circuit of the amplifier as is illustrated. For example, if the filament voltage increases, cathode emission of both tubes will increase. However, the increased plate current of the diode must pass through the cathode resistor of the amplifier. This increased component of cathode resistor current results in an increase in bias which tends to nullify the effects of the original increase in filament voltage. Duplex tubes with diode sections having the same filament as the triode or pentode sections are especially useful in compensation circuits of this type.

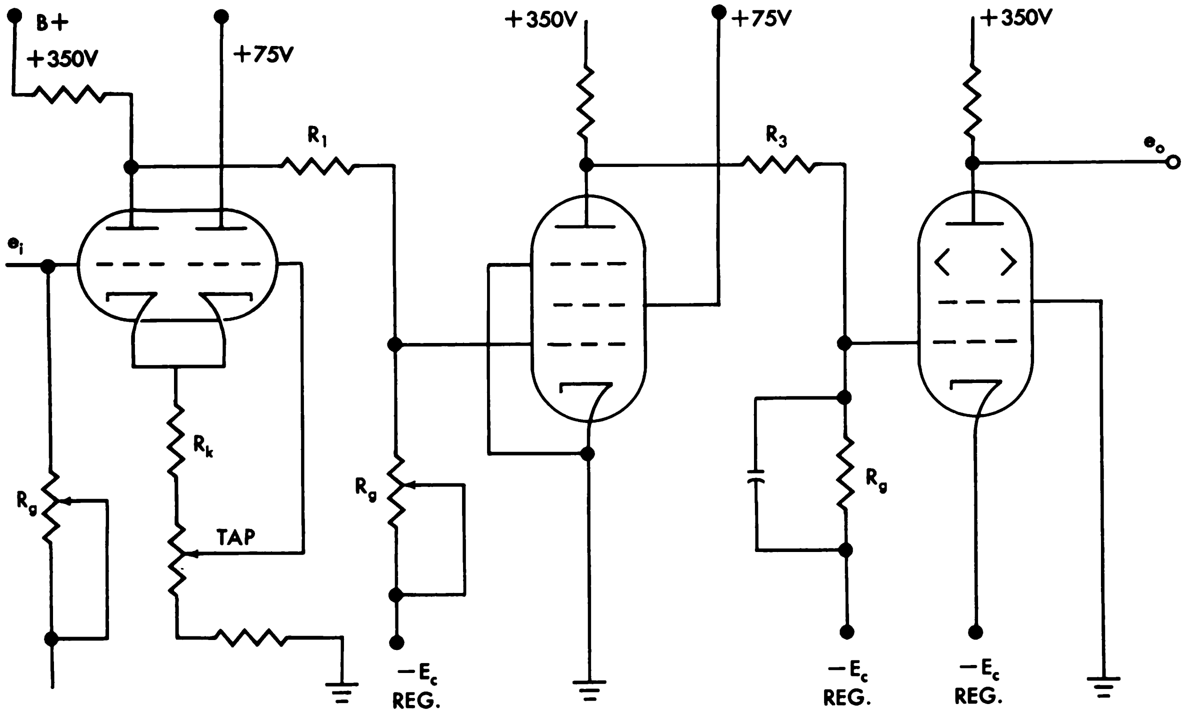


Diode Compensator

UNBALANCE DUE TO GRID CURRENT. Besides drift, there may be an undesirable component of output voltage due to grid current. Currents flowing into or from the grid cause undesirable voltage drops across the grid resistors, leading to distortion errors in the computer variables. For example, if the input grid level is to be maintained at zero with zero input and there happens to be only .001 microampere flowing through a 10 megohm grid resistor, an error voltage of .01 volt results. Hence, it is obvious that grid currents may lead to serious computing errors when their effects are amplified and/or integrated.

The grids of some tubes may even emit electrons if some of the oxide from the cathode is deposited on the grid. Also, if the grid is maintained at a high negative value, it may attract positive ions prevailing in the residual gas of the vacuum tube. It is usually helpful to operate DC computer amplifier tubes at low values of plate, screen, and filament voltages in order to minimize ionization. The usual practical remedy for troubles due to grid current is to use vacuum tubes whose inherent grid current is sufficiently small. Fortunately, vacuum tubes of the electrometer type are available; they suit most requirements.

A Typical Computer Amplifier utilized in a contemporary fire control computer is illustrated next. In this amplifier high voltage gain is obtained without the use of regeneration through the use of a pentode in the second stage and a beam power tube in the output stage. A differential amplifier is used in the input stage in order to compensate for the effects of cathode mission fluc-



A Typical Computer Amplifier

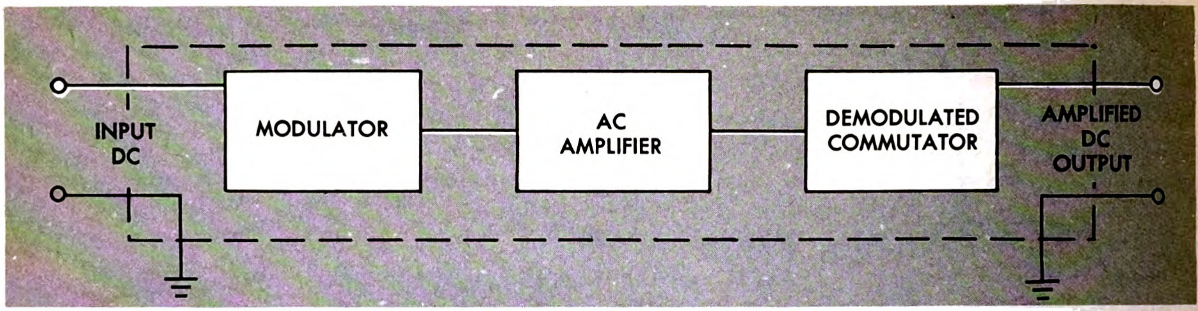
tuations. In order to minimize computer errors due to drift, the amplifier should be balanced before each computer operation. In order to prolong the effectiveness of compensation, all tubes should be aged by operating them through extremes of heater voltage. After proper aging, the cathode resistor tap of the first stage should be adjusted so that the portion of R_k between the tap and ground is as nearly as possible equal to r_p/μ, where r_p is the dynamic AC plate resistance of either tube and μ is the amplification factor.

COMPONENTS OF MODULATED CARRIER TYPE COMPUTER AMPLIFIERS

Modulated carrier type computer amplifiers are frequently used in connection with various fire control, guidance, and attitude computer systems where balancing and zero adjustment must be accomplished automatically. A block diagram illustrating the principle of a modulated carrier type computer amplifier is shown. This unit is sometimes referred to as a *chopper and synchro amplifier.*

The DC voltage applied to the input stage amplitude-modulates the output of a stable continuous wave synchronous vibrator. That is, the AC amplitude is made to vary in accordance with the magnitude of the input DC. The modulated AC is coupled, as shown, to a conventional AC amplifier. The amplitude-modulated carrier signal is then demodulated (rectified and filtered) by some type of commutator device to produce an amplified DC proportional to that originally applied at the input.

A modulated carrier type amplifier will be discussed in more detail later when investigating the principles of automatic drift control and balancing circuits. At this point it may be well to investigate briefly the various types of modulators and demodulators utilized in computer networks.



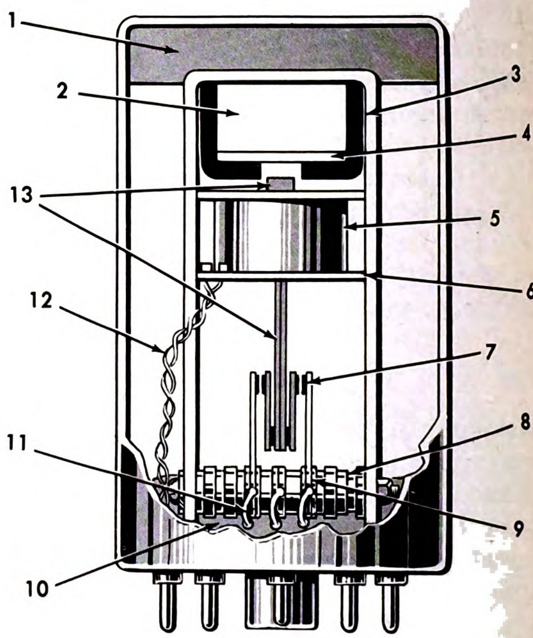
Principle of a Modulated Carrier Computer Amplifier Used in Automatic Balancing Circuits

Modulators and Demodulators

From the previous illustration we can conclude that electrical signals in computer amplifiers may be either AC or DC. Since some of the computer components may require DC and other components in the same loop may require AC, it is desirable to have devices which can either convert AC signals to DC, or DC signals to AC.

In a computer system, modulators are generally used to convert polarized DC signals to properly phased AC signals whose amplitude varies in accordance with the magnitude of the input DC. A demodulator, sometimes referred to as a *commutator,* accomplishes a conversion similar to that of a modulator, but in the reverse order. It receives an AC input of a particular phase and produces a polarized DC output. There are several types of modulators and demodulators utilized in interceptor computer systems which will perform the functions cited above. We will investigate a few more common types of mechanical and vacuum tube modulators and demodulators.

MECHANICAL CHOPPER MODULATOR. A mechanical chopper is essentially an AC driven relay or synchronous vibrator that may perform switching operations at frequencies up to about 800 c.p.s. A cutaway view of a mechanical chopper of the synchronous vibrator type is shown in the accompanying illustration. Note that this particular chopper looks like a vacuum tube, having a similar base suitable for plugging into a similar



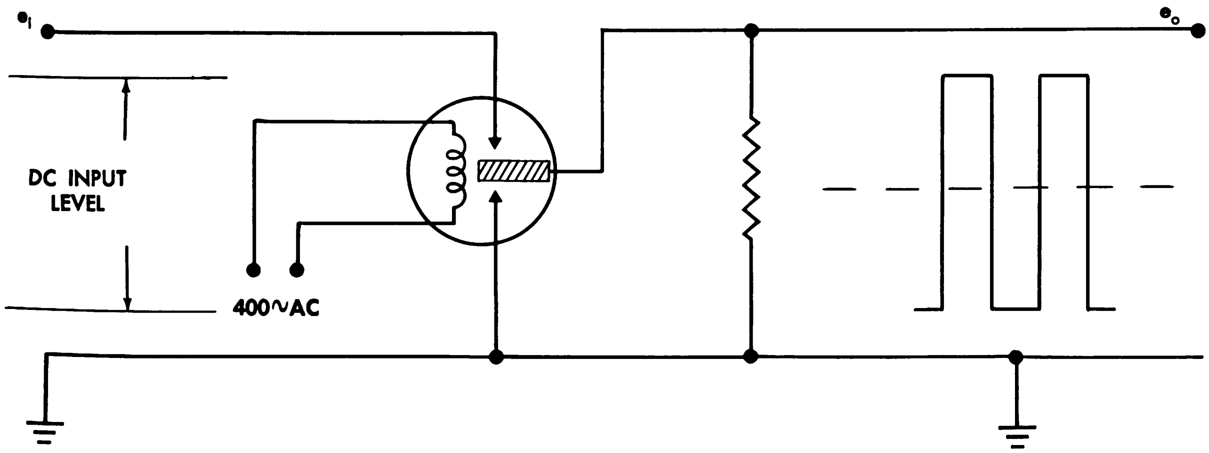
1. SILICON SPONGE CUSHION
2. PERMANENT MAGNET
3. SOFT IRON
4. BRASS PLATE
5. COIL
6. FIBRE SPACERS
7. POINTS
8. METAL SPACER
9. INSULATION SPACER
10. CUSHION
11. ELECTRICAL LEADS
12. ELECTRICAL LEADS
13. VIBRATING REED

Cutaway View of Mechanical Chopper

socket. The root of the vibrator reed serves as the core for the AC excitation coil. This excitation winding is energized by AC of predetermined frequency (usually 400 c.p.s.). Thus, its core becomes an electromagnet whose upper end changes polarity on alternate half-cycles. The root of the vibrator reed is, therefore, attracted to opposite ends of the permanent magnet on alternate half-cycles, and the vibrator reed makes and breaks contact with the vibrator points at the rate of 400 c.p.s.

When a synchronous vibrator of this type is used as a modulator, the vibrator reed may be connected to the output terminal as shown in the circuit diagram. Also, one of the contact points may be connected to the input and the other to ground. In this manner the output line is connected alternately to the input DC level and to ground or zero level DC. Hence, the vibrator reed intercepts the output DC level, resulting in a square wave having the level of the DC input for one half-cycle and having a zero level for the other half-cycle. The blocking condenser following the output resistor removes the DC component, so that the signal becomes an AC square wave with an amplitude proportional to the DC input level and a phase which reverses when the DC input changes polarity.

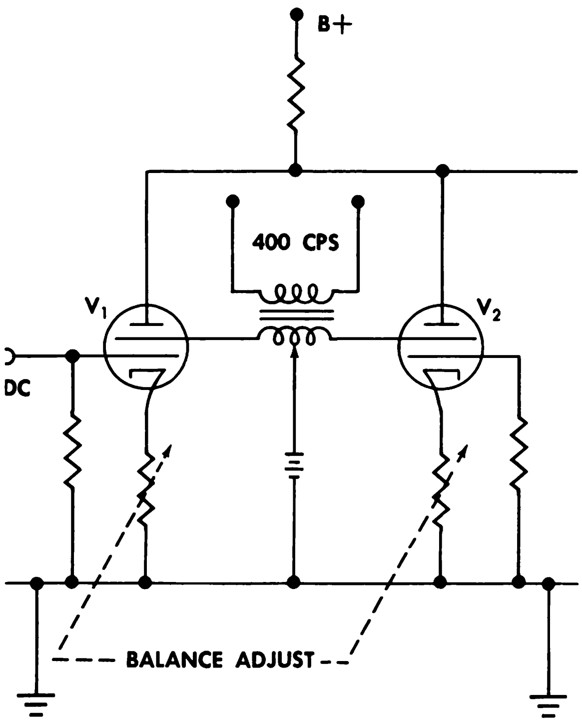
The main source of error voltages in synchronous vibrators is due to stray coupling of the AC synchronous driving voltage into the carrier signal circuit. This effect is, of course, minimized by proper shielding and phasing of the synchronizing voltage. In some applications, the driving coil terminals are brought out through the top of the vibrator case well away from the signal carrying leads. It may be well to point out that it is often found desirable to make the contact reed of the vibrator vibrate at a frequency equal to twice the excitation frequency applied to the synchronizing coil so as to further reduce the effects of AC pickup at the exciting frequency.



Mechanical Chopper Modulator

VACUUM TUBE CHOPPER MODULATOR. The next figure illustrates a typical vacuum tube chopper modulator circuit. In this circuit the voltage gain or the amplitude of the AC component of output voltage is modulated by the level of the DC voltage left on the control grid of V_1. If the control grid of V_1 is made less negative with respect to its cathode, the gain and therefore the amplitude of the AC component of plate voltage will be increased. Alternating voltage of the desired frequency is obtained by exciting the screen grids of both tubes through opposite ends of the transformer secondary. Note that the center tap of the secondary winding is tied

to ground through $B+$, thus providing the proper average DC level (positive with respect to ground) on both screen grids.



Vacuum Tube Chopper Modulator

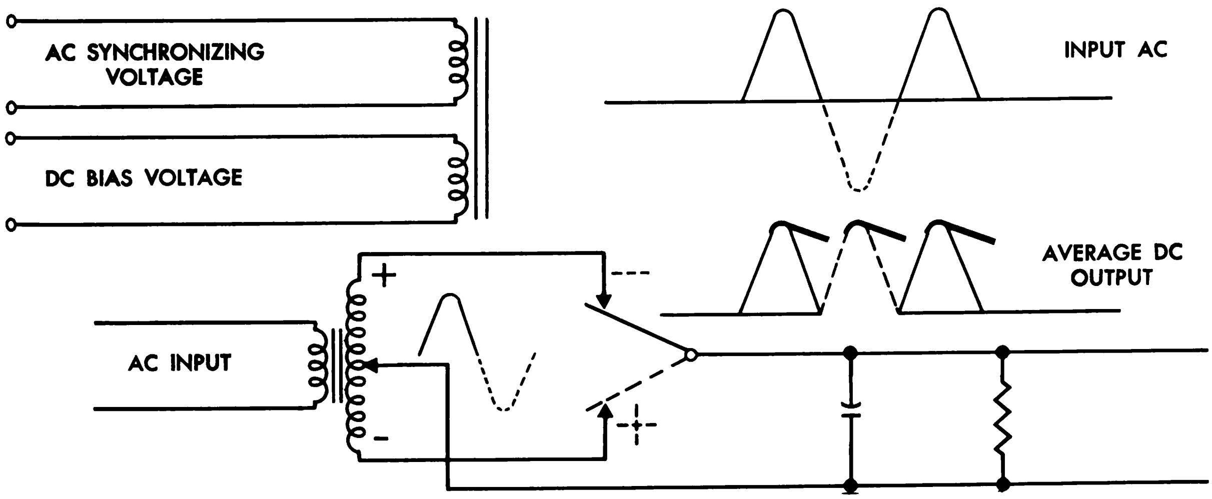
In the absence of a DC input signal, the output voltage should also be adjusted for zero AC. This may be accomplished by adjusting the variable cathode resistors so as to produce equal bias and equal voltage gain for both tubes. As a result, both tubes conduct equal DC components of current through the common plate load. Since the AC components of voltage on the two screen grids are made 180 degrees out-of-phase with each other, the effect of raising the voltage on one screen is exactly balanced by that of lowering the voltage on the other screen. Consequently, the total current flowing through the common plate load is unchanged, and there can be no AC output.

If a DC is applied to the input of V_1, the bias level of V_1 and V_2 become unequal. This results in unequal DC components of plate current as well as an unequal amplification of the alternating components. Thus, the AC components of plate current resulting from the screen grid excitation voltage will no longer balance out. In fact, there will now be an alternating current flowing through the common plate load which is equal to the difference between the AC plate currents of each tube. An AC will, therefore, appear at the output terminals with an amplitude proportional to the input DC level and a phase which reverses when the input DC polarity is changed.

MECHANICAL DEMODULATOR. A demodulator or commutator of this type is also known as a *phase sensitive detector*. As previously mentioned this type modulator makes a conversion similar to that of a modulator, but in reverse order. It receives an AC input and develops a DC output. For an input of one phase, it develops a positive DC output. For an input with a phase difference of 180 degrees, it produces a negative DC output.

Mechanical demodulation may be accomplished with a synchronous vibrator similar to the permanent magnet type previously treated in conjunction with mechanical modulation. It may be accomplished with the use of a synchronous mechanical switch or relay such as is illustrated in the schematic diagram.

The Synchronous Mechanical Switch depends on a vibrating arm similar to the vibrator reed of the mechanical chopper. Note, however, that the vibrating arm does not act as the core for the AC excitation winding, but as a spring-loaded armature which is attracted to a magnetized core. The excitation coil is energized by a synchronizing AC of the desired frequency. DC current is passed through the bias coil (wound on the same core) in order to prevent the vibrating arm from being attracted during both the positive and negative half-cycles. That is, when the synchronizing AC flows in one direction, its magnetizing force is *aided* by that of the DC and the vibrating arm is attracted. But when the AC flows in the



Mechanical Demodulation with Synchronous Switch or Relay

opposite direction, its magnetizing force is opposed by that of the DC and the attraction of the arm is reduced. The arm, thus, vibrates at the same frequency as the synchronous AC if the DC bias coil is also energized. If the bias coil is not energized, the chopper relay becomes a nonpolarized type, so that the armature is attracted on both negative and positive half-cycles, providing a carrier frequency which is double that of the synchronizing frequency.

Now let us assume that an AC signal (whose frequency is equal to the synchronizing frequency) is to be demodulated by the synchronous mechanical switch shown above. Note that the signal to be demodulated is fed to the alternate relay contacts through the opposite ends of a transformer secondary whose center tap is connected to a common ground. During the positive half-cycles (as indicated by the solid waveform and polarity markings) the upper contact is positive with respect to ground, and the vibrating arm is up. Thus, a positive pulse is produced across the low-pass filter. During the negative half-cycles the arm is down; but the lower contact is then positive with respect to ground, and positive pulses are again produced across the output terminals.

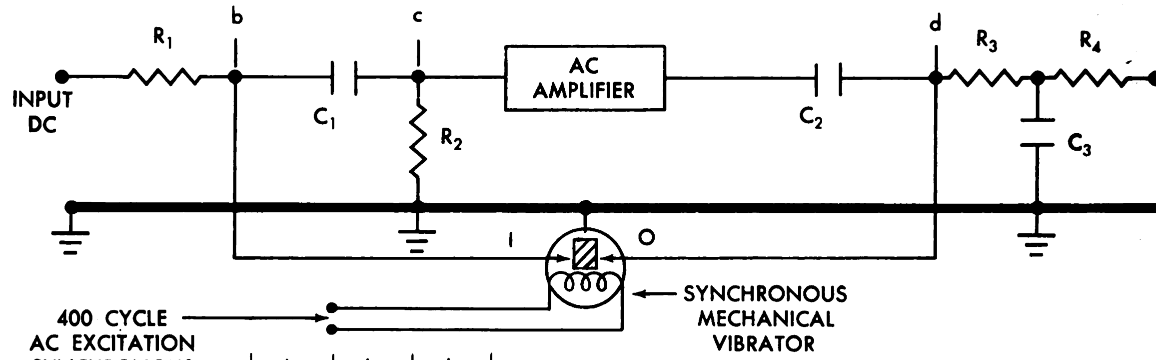
The resulting output is similar to that of a full-wave rectifier. After being filtered, the output is a DC voltage whose magnitude is proportional to amplitude of the AC input.

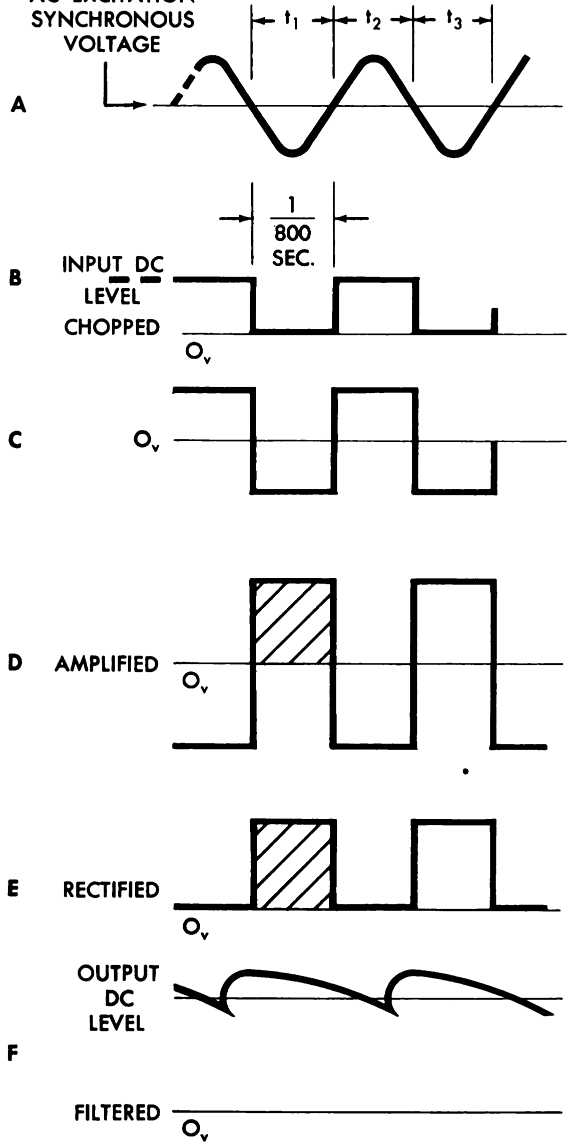
MODULATED CARRIER D.C. AMPLIFIER

Modulated carrier type amplifiers permit a high degree of DC amplification with very low drift and without the use of carefully regulated power supplies. The figure and accompanying waveforms illustrate the use of a single mechanical chopper or synchronous vibrator to perform both modulation and demodulation. You will find, upon investigation, that the DC input voltage modulates a 400-cycle AC carrier by the use of the chopper contact on the left to interrupt the DC input level. The resulting amplified modulated carrier is then rectified synchronously by the contact on the right of the same chopper.

Modulating the Synchronous Carrier

In the circuit illustrated, the center contact of the synchronous vibrator is grounded. The left contact, designated as the input contact *I*, is connected to the input (terminal *b*) of the AC amplifier through the DC blocking capacitor C_1. The contact on the right, designated as the output contact *O*, is connected to the output plate of the AC





Modulated Carrier Type Computer and Its Principle Wave Forms

amplifier through the DC blocking capacitor C_2. The grounded vibrator reed makes and breaks contact with the contacts *I* and *O* at the frequency of the excitation carrier signal.

When the chopper is energized with a 400-cycle AC (as indicated at A in the illustration) the grounded reed vibrates at this frequency. During the negative half-cycles, the input contact *I* is grounded through the grounded vibrator reed, thus shorting out the input signal and reducing the DC input level to zero for a period of 1/800 second. During the positive half-cycles of the synchronizing signal, the input contact is ungrounded, allowing the input to be restored to its original level. Hence, the DC input level is interrupted periodically at the rate of 400 times per second, producing a square waveform of voltage which alternates between zero and the initial voltage level as shown at B in the illustration. This chopped input (square wave) is coupled to the grid of a conventional AC amplifier through a long time constant network consisting of R_2 and C_1.

Amplifying the Modulated Carrier

The DC component of the alternating carrier (square wave) is blocked by C_1, permitting only the AC component (shown at C) to be fed to the input of the AC amplifier. The waveform shown at D illustrates the amplified version of the chopped DC input. This waveform appears at point d in the circuit, before it is itself rectified, and assumes the amplified signal to be distortion-

less. However, there will be some change of the waveshape due to attenuation of the higher frequency components and some phase shift, both of which are neglected here for simplicity.

Demodulating the Modulated Carrier

During the positive half-cycles of the 400-cycle synchronizing signal, the output contact O is grounded through the vibrating reed, thus shorting to ground the output plate voltage of the AC amplifier, which reduces the output at point d to zero during period designated as t_2. During the period t_3 the output terminal is again ungrounded, permitting the output plate voltage at point d to be restored to its maximum amplified level. The waveform illustrated at **E** then represents the rectified but as yet unfiltered, amplified version of the chopped DC input signal. This pulsating DC is then smoothed by the filtering network composed of R_3, R_4, and C_3. The resulting amplified DC proportional to the instantaneous DC input level is shown at **F** in the illustration.

It may be well to point out that this modulated carrier amplifier is phase sensitive. That is, if the sign of the DC input voltage is reversed, the phase of the modulated AC also changes by 180 degrees, and the sign of the output DC will reverse also.

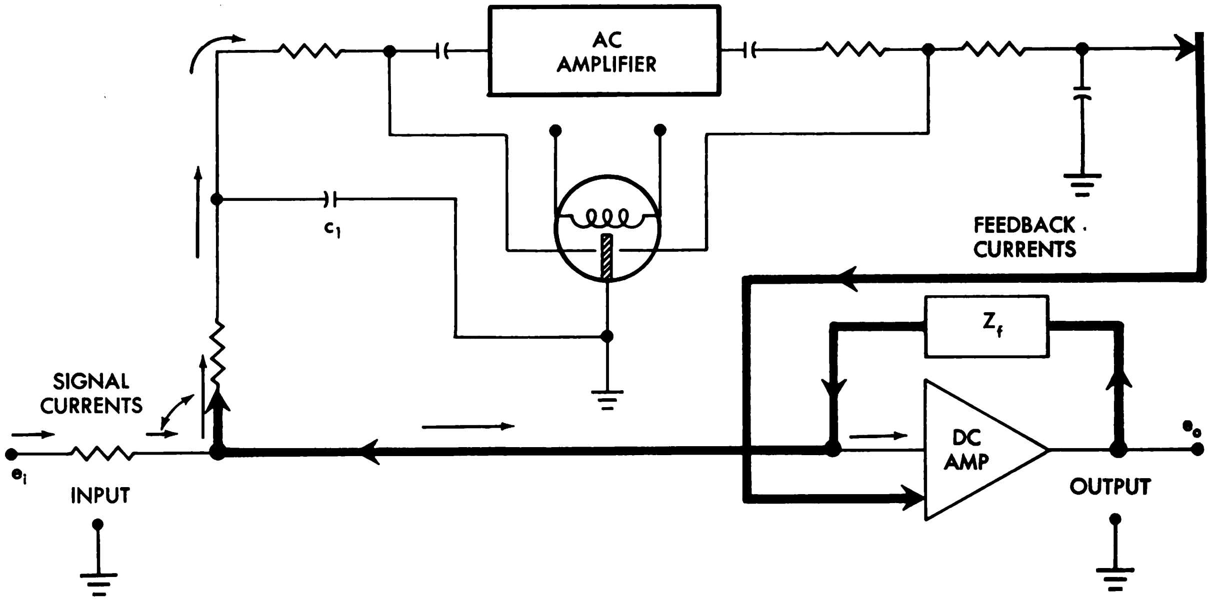
Limits of Operation

Mechanical choppers are commonly operated at frequencies of about 200 c.p.s. and seldom operated at frequencies higher than 400 c.p.s. You can see that it would be impossible to properly modulate a 200-cycle carrier frequency with a DC input level that is varying at a higher rate. Hence, it is not possible to employ a modulated carrier type amplifier using a mechanical chopper directly as the main amplifier in high-speed computing applications where the input DC level may vary at higher frequencies. In fact, modulated carrier type amplifiers are limited to a rather low rate of change in DC input levels. They are restricted to variations in input levels well below the synchronous carrier frequency of the mechanical chopper involved. If a modulated carrier type amplifier is to be used (alone) as an operational DC amplifier, it is desired that the input voltages representing variables do not vary significantly during the modulation period.

Higher carrier frequencies may be obtained by replacing the mechanical chopper with diode or triode type switching circuits. However, the drift inherent in diode and triode modulators and demodulators is of the same order as the basic DC amplifiers previously discussed.

The same bandwidth resulting from the low-pass filters required is another objection to the use of a modulated carrier amplifier (alone) as the main amplifier. Dynamic problems often include motions or other variables that are represented by DC voltages which vary at frequencies up to about 15 c.p.s. and the low-pass filters will not respond quickly enough to changes in amplitude of the resulting modulated carrier. However, as far as drift correction is concerned, this restriction is unimportant because drifts such as variations in cathode emission, grid current, and power supply voltages occur very gradually. Consequently, the modulated carrier type amplifier may be used very advantageously as additional equipment to reduce drift in DC computer amplifiers.

COMBINING A MODULATED CARRIER AMPLIFIER WITH A CONVENTIONAL DC AMPLIFIER. In the accompanying block diagram a modulated carrier amplifier is combined with a conventional DC amplifier to form an *Automatic Drift Corrected and Balanced Computer Amplifier.* In this arrangement the error due to drift, appearing at the output terminal of the DC amplifier, is fed back through the feedback network. That is, the error voltage is fed back to the input of the modulated carrier channel, where it tends to charge C_1 of the lowpass filter. This error voltage is, in turn, amplified by the modulated carrier amplifier and again fed back



Automatic Drift Corrected and Balanced Computer Amplifier

in such a manner as to nullify the effects of the original error voltage due to drift. The lowpass filter—RC network at the input of the modulated carrier channel—senses the errors due to drift, accumulates and remembers the effects of past disturbances for a short time, and seeks to nullify their effects. You may recall that this is a type of subtractive integral control discussed in the section on integrating components of computers.

Although modulated carrier type DC amplifiers can be engineered to give good performance, the relay itself is not completely satisfactory because its working life is limited. It is, then, natural that attempts have been made to replace mechanical choppers with some nonmechanical device. Modulators and demodulators using diodes and triodes may be used, but, as previously mentioned, they are subject to drift errors of the same type as they are intended to avoid. Vacuum tubes consume a good deal of energy in the form of filament dissipation, they are fragile, and their life is also limited. For these reasons, the elimination of vacuum tubes from computing equipment seems to be almost universally favored.

Present work in progress seems to indicate that semiconducting solid state devices may replace the vacuum tube. There is, at least, great hope that performance of drift corrected amplifiers will be improved by using germanium diodes, transistors, and saturable reactors as modulators and demodulators. A great deal of development and vigorous progress is currently taking place in this field. The remaining portions of this section will be devoted to a study of such solid state devices as transistors and magnetic amplifiers.

TRANSISTORS

Contemporary interceptor aircraft and missiles are highly complicated mechanisms, and every attempt is made to relieve the pilot and crew members of tasks which can be done by machines. This practice, although advantageous, requires the use of more and more intricate mechanisms. As

their complexity increases, the problem of reliability becomes more acute. In recent years, the advancement and refinements in electronic automatic control equipment has been remarkable. These automatic features have resulted in simplified operation but, at the same time, have resulted in the need for extremely complex circuitry.

The current trend in interceptor aircraft design seems to be toward smaller aircraft with more electronic equipment in them. Subminiaturization is, therefore, becoming a very necessary requirement. One of the most important developments in the field of electronics, aiding in the trend toward miniaturization, is known as the transistor. The transistor, the name of which was coined from TRANSient and resISTOR, is a miniature substitute for the vacuum tube. The transistor is capable not only of doing most of the things that vacuum tubes do but can also perform jobs which have heretofore been difficult, if not impossible, for vacuum tubes alone to handle.

Transistors may be utilized as synchronous electrical choppers for converting weak DC input signals into proportional square wave AC signals. The modulated AC signal may then be amplified by transistorized AC amplifiers and, in turn, be demodulated by transistor commutators. Hence, it is possible to completely transistorize low-level DC amplifiers which will operate over an extremely wide temperature range without compensation and yet have only slight drift due to temperature changes. Other than being used in chopper modulators and demodulators for DC amplifiers, the transistor may be used as an oscillator for telemetering service, as a diode or triode detector, as an impedance-matching device, and as a mixer. Along with its amazing versatility, the transistor is much smaller, lighter, longer-lived, more rugged, more efficient, and potentially less costly than the vacuum tube. The transistor offers the additional advantage of instant operation; that is, no warmup time is required nor is standby power needed for its operation. The size and the fantastically small power requirements of transistor circuits make them ideal for portable equipment, remote control circuits, and wherever light weight, small size, and economy of operation are desirable in electronic equipment.

Electron Theory of Transistors

The operation of transistors depends on the electrical properties of a class of substances known as semiconductor crystals. A semi-conductor is a combination of crystals in which, under certain conditions, the combination acts like a conductor but, under other conditions, acts like an insulator. Almost all semiconductors permit electrical current to flow more readily in one direction than in the other and, therefore, may be used as diodes or rectifiers.

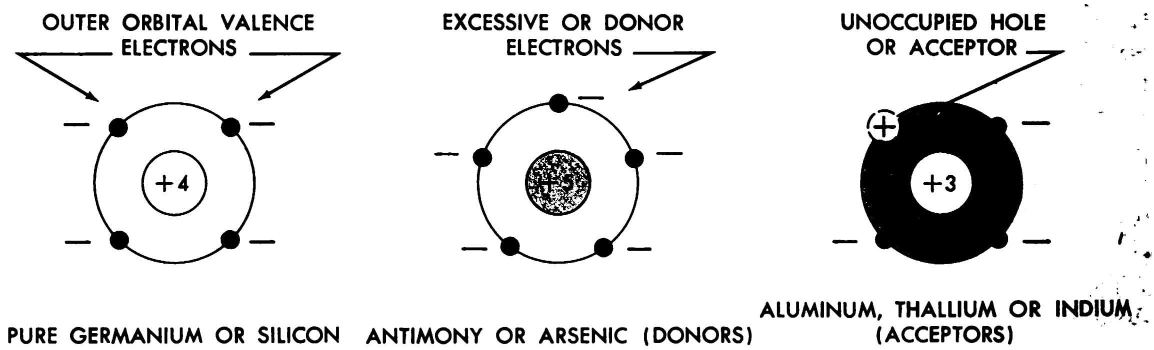
Selenium, silicon and germanium, when combined with certain impurities, are the most popular semiconductors. There are other materials exhibiting similar properties, but they are not used as extensively as the three just mentioned. Selenium is widely used in power rectifiers and in self-generating photo cells. Silicon is used in high-frequency mixer diodes. Germanium is used in diodes, transistors, and phototransistors.

As a step toward understanding the electron theory of transistors, you will find it helpful to first consider the atomic structure of atoms.

Structure of Atoms

As you may know, all matter is composed of one or more elements, and each element is composed of atoms. The atom, in turn, is composed of smaller units of matter called electrons, protons, and neutrons. Electrons possess a negative charge. Protons possess an equivalent positive charge, and each is more than 1800 times as heavy as an electron. The neutron possesses no charge and has the same mass as the proton.

The atom, which is the smallest subdivision of an element possessing all the



Symbolic Atomic Representation of Various Atoms

properties of the element, is composed of a nucleus of protons and neutrons around which the lighter electrons revolve. Atoms of one chemical element differ from those of another element only in the number of electrons, protons, and neutrons in their structure.

In the normal atom there are as many electrons—negative charges—outside the nucleus as there are protons—positive charges—within the nucleus. The study of atomic structure has shown that the inner orbital electrons in the lower energy levels around the nucleus are tightly bound to it and do not enter in to chemical reactions or transistor physics. Electrons in the highest energy levels, in the outermost orbit of atoms, are those which cause chemical combinations and are referred to as valence electrons. The nucleus and the tightly bound electrons comprise an inert core with a net positive charge around which the less tightly bound valence electrons revolve.

GERMANIUM ATOM. In transistor physics we will be concerned with the net charge on the core and the electrons surrounding it. For example, each atom of pure germanium has 32 protons in its nucleus. It has 28 tightly bound inner orbital electrons and 4 relatively free valence electrons in its outermost orbit. Hence, pure atoms of germanium may be represented by a core with a net charge of $+4$ surrounded by four valence electrons. Each atom of silicon has 14 protons in its nucleus and 10 tightly bound inner orbitual electrons. Hence, an atom of silicon may be symbolically represented in the same manner as an atom of pure germanium as shown at the left in the illustration.

DONOR ATOMS. Antimony and arsenic atoms become *donors* when they are combined with the crystal structure of otherwise pure germanium. The net charge on the cores of antimony and arsenic atoms is $+5$. Each atom has five valence electrons surrounding its core as indicated at the center.

ACCEPTOR ATOMS. An *acceptor* atom may be defined as an atom which, when combined with the crystal structure of otherwise pure germanium, produces a *hole,* or an excess positive charge within the crystal. A *hole* may be defined as the unoccupied space of a required valence electron, resulting in an incomplete group of electron-pair valence bonds whose conduction properties are similar to those of an electron except that it carries a positive charge instead of a negative one.

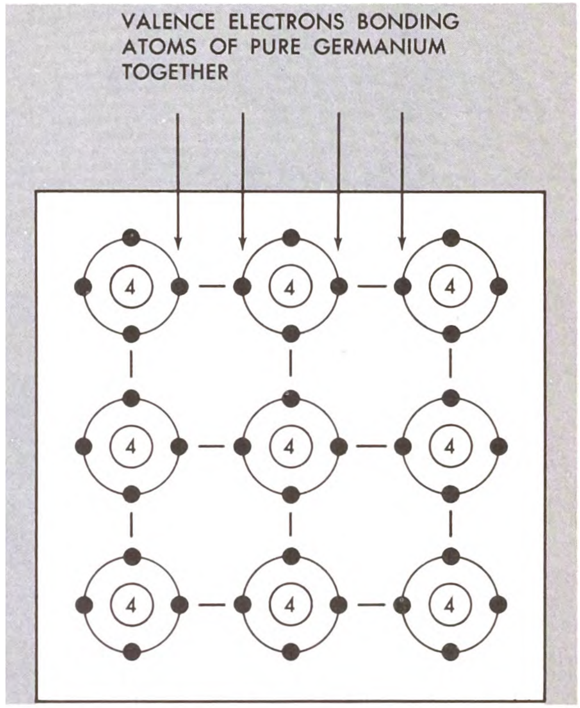
Aluminum, thallium, and indium atoms become *acceptors* when they are combined with the crystal structure of otherwise pure germanium. The net charge on the inert cores of their atoms is $+3$, and each of these atoms has three valence electrons surrounding its core. The symbolic atomic representa-

tion of these acceptor atoms is shown at the right in the illustration.

Structure of Germanium Crystals

The molecular theory of crystals has shown that the atoms of pure germanium form themselves into distinct, characteristic, crystalline shapes and that the atoms in the crystal lattice are held together by their loosely bound valence electrons which adjacent atoms share. As indicated in the diagram illustrating *A Portion of a Molecule or Crystal of Pure Germanium,* each atom has four neighbors which are equidistant from each other. Between the cores of the atoms and each of their neighbors are two electrons. These electron pairs form electron-pair valence bonds which come into existence when two or more atoms approach each other; that is, the movement of an electron from one atom becomes coordinated with the movement of electrons from its neighboring atoms. At low temperatures, the coulomb forces of repulsion and attraction seem to be in perfect balance, and the atoms are said to be in a condition of equilibrium. Hence, the loosely bound electrons are held by the bonding energy of the valence bonds and are not available for current flow. Consequently, the pure germanium acts like an insulator having a high resistance. At higher temperatures, thermal agitation causes the valence bonds to lose effect, and electrons are made available for current flow. Thus, pure germanium crystals have the properties of a semiconductor with a negative temperature coefficient. This is one type of transient resistance; that is, the resistance varies inversely as the temperature.

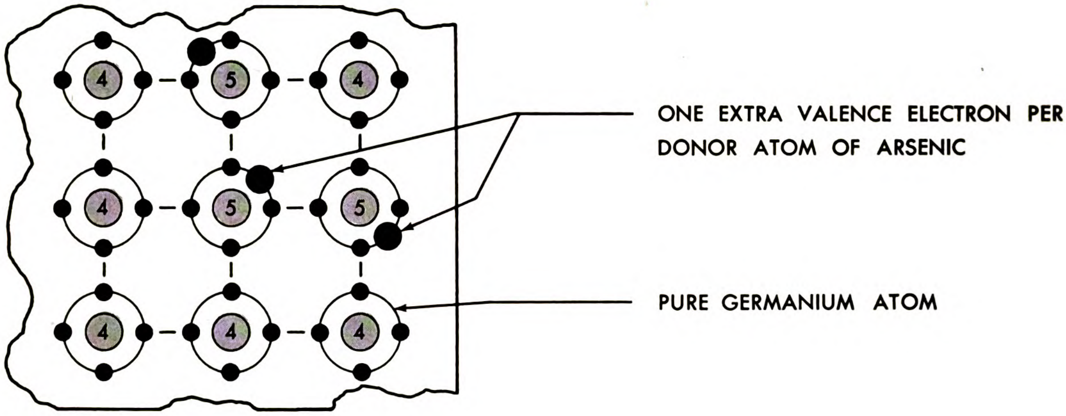
We are concerned, however, with semiconductor crystals whose electrons need little energy, room temperature only, for continuous displacement and recombination. Under some conditions free electrons may be detached from the atoms of a crystal and made to travel through its substance much in the same manner in which normal current flows through an ordinary conductor.



A Portion of a Molecule or Crystal of Pure Germanium

Those crystals which permit current flow primarily by means of free electrons are called *N* or negative type crystals.

N-TYPE GERMANIUM CRYSTAL. A typical atomic diagram of an N-type germanium crystal is illustrated. Note that the crystal illustrated is composed of pure germanium with a small amount of arsenic added as a donor impurity. You will recall that an atom of arsenic contains 5 electrons in its outer orbit. Hence, it has one more than the required number of valence electrons per atom as indicated in the illustration. When a donor atom of arsenic is added to the crystal structure of germanium, the donor is free to lose one of the five electrons. That is, one of these electrons is made available for current flow because only four of them can form electron-pair valence bonds with the electrons of its neighboring germanium atoms, thus allowing the excess electron to move through the relatively wide spaces between the cores.

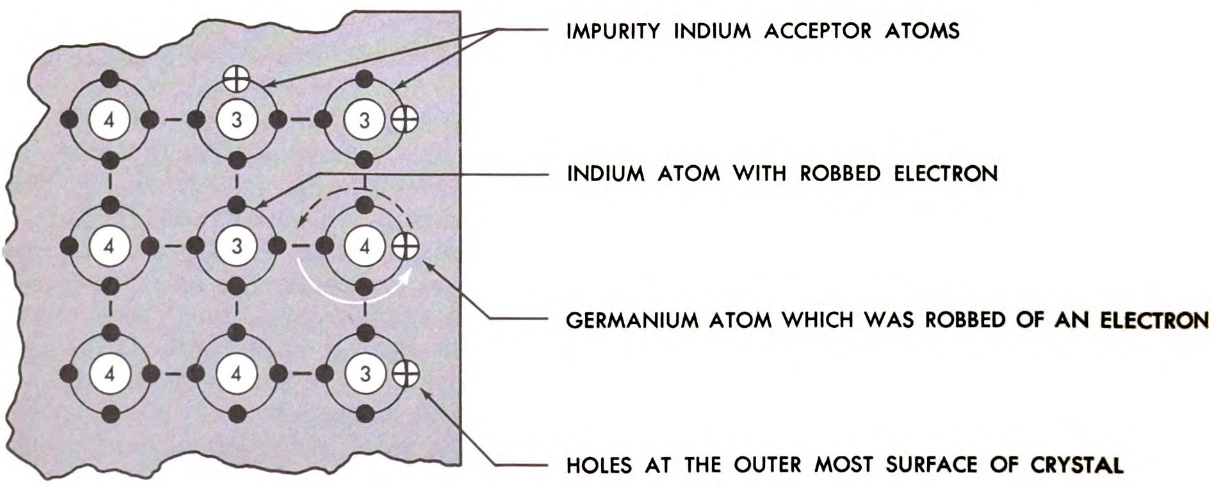


N-Type Germanium Crystal

If a battery were placed across the crystal, the free (excess) electrons from the donor atoms would move toward its positive terminal and enter the battery at that point. At the same time, an electron would leave the crystal. Hence, a continuous electron displacement and recombination (current flow) would take place, but the cores of the germanium and donor atoms would remain undisturbed.

P-TYPE GERMANIUM CRYSTAL. Those crystals which permit current flow primarily by means of *holes* are called *P* or positive type crystals. A typical atomic diagram of a *P-Type Germanium Crystal* is **illustrated.** Note that this crystal is composed of pure germanium with a small amount of indium added as an *acceptor impurity.* Remember that an indium atom contains only 3 electrons in its outer orbit; hence, it has one less than the required number of valence electrons per atom, creating a hole as indicated in the illustration.

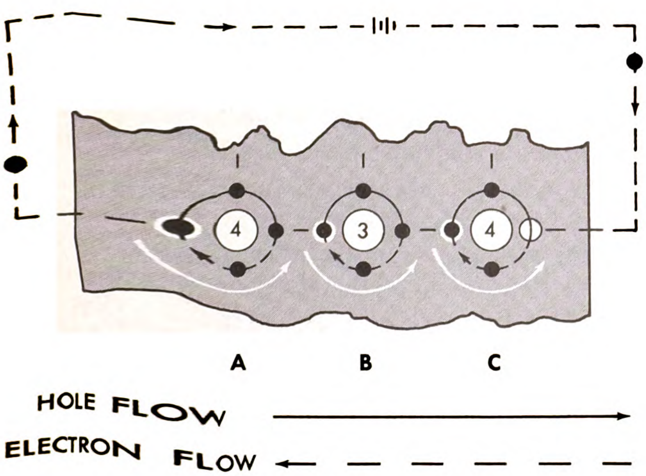
Since an acceptor atom of indium has a hole—an inadequate number of valence electrons to form the required number of electron-pair valence bonds—it may combine with the crystalline structure of germanium



P-Type Germanium Crystal

by accepting or robbing an electron from one of its neighboring germanium atoms, thus forming the required number of electron-pair valence bonds to bind the dissimilar atoms. However, when this new electron-pair bond is formed, a hole is created in the neighboring germanium atom which was robbed of an electron. This displacement and recombination of electrons and holes continues until the holes have migrated to the indium and/or to the germanium atoms at the outer surfaces where valence bonding electrons are not required. At the same time the electrons migrate and fill the holes in the inner regions where electron-pair bonds are required. Thus, the P-type germanium crystal is formed as illustrated. In the illustration, migration of electrons is indicated by dashed lines and migration of holes is indicated by solid lines.

Conduction of Holes and Electrons in a P-Type Germanium Crystal is illustrated in the next figure which considers only three horizontal atoms of the previous illustration. At the instant the battery is placed across the crystal, the hole in the atom at *C* will be attracted toward the negative terminal of the battery, and an electron from this terminal will enter the crystal and fill the hole.



Conduction of Holes and Electrons

Simultaneously, an electron from atom *A*, which is near the surface at the positive terminal of the battery, will separate from its bond and enter the battery, thus creating a hole in atom *A*. This hole in atom *A* will then rob an electron from atom *B*, leaving *B* with a hole and neutralizing *A*. In turn, the hole created in *B* will rob an electron from *C*, leaving *C* with the hole.

Hence, the hole that is originally created at *A* migrates to *C*, where it is again attracted toward the negative terminal of the battery and allowed to accept another electron from the source of supply. This sequence of displacement and recombination events is repeated so as to maintain continuous current (holes and electrons) through the crystal. Note that holes may thus migrate through a P-type germanium crystal acting, for all practical purposes, like current flow of positive charged particles.

Regardless of whether a particular substance is an N- or P-type material, it must be remembered that *holes* may exist in the N-type material and that free electrons may exist in the P-type material. It must also be kept in mind that *both holes and free electrons may be formed by passing current through a material.*

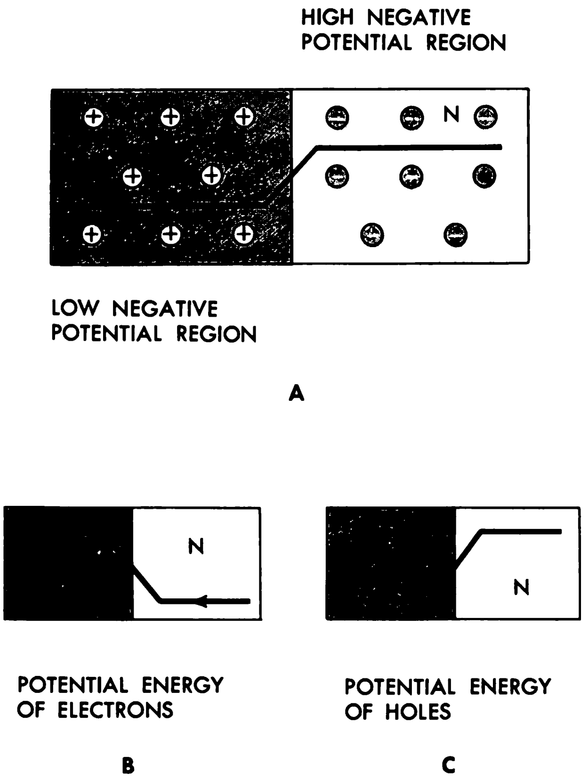
Germanium Diode Semiconductors

A semiconductor is a combination of crystals in which, under certain conditions, the combination acts like a conductor, but, under other conditions, acts like an insulator. Almost all semiconductors permit electrical current to flow in one direction more readily than in the other and therefore may be used as diodes or rectifiers.

You must bear in mind, however, that the individual germanium crystals themselves are good conductors and can conduct equally well in either direction. Rectification with germanium crystals may occur if P-type germanium and N-type germanium are placed side by side. The plane at which the two types of germanium meet is called a *P-N junction,* and the action which occurs

at the junction constitutes the basic action of transistor operation.

P-N JUNCTION. The figure illustrates a P-N type junction in the state of equilibrium. Note, at A, that only holes and electrons of the acceptor and donor atoms are shown—holes remaining in the *P-region of lowest negative (electrostatic) potential* and electrons remaining in the *N-region of highest negative potential.* As shown in the accompanying *potential energy curves* at B, the electrons remaining in the region of highest electrostatic potential have a minimum of potential energy since potential energy means ability to do work. Since an electron cannot move to do work after it reaches the region of highest negative potential, it becomes apparent that the electron in this N-region has low potential energy. Hence, the N-region is the low potential energy region for the electrons.



P-N Junction in State of Equilibrium and Potential Energy Curves

Bear the preceding statement in mind because potential energy diagrams are used often in illustrating transistor behavior.

The same statement can be applied with reference to holes, for, when the hole is in the region of lowest electrostatic potential, its potential energy is at a minimum as illustrated at C. Note that the low potential energy region for electrons is a high potential energy region for holes, and vice versa.

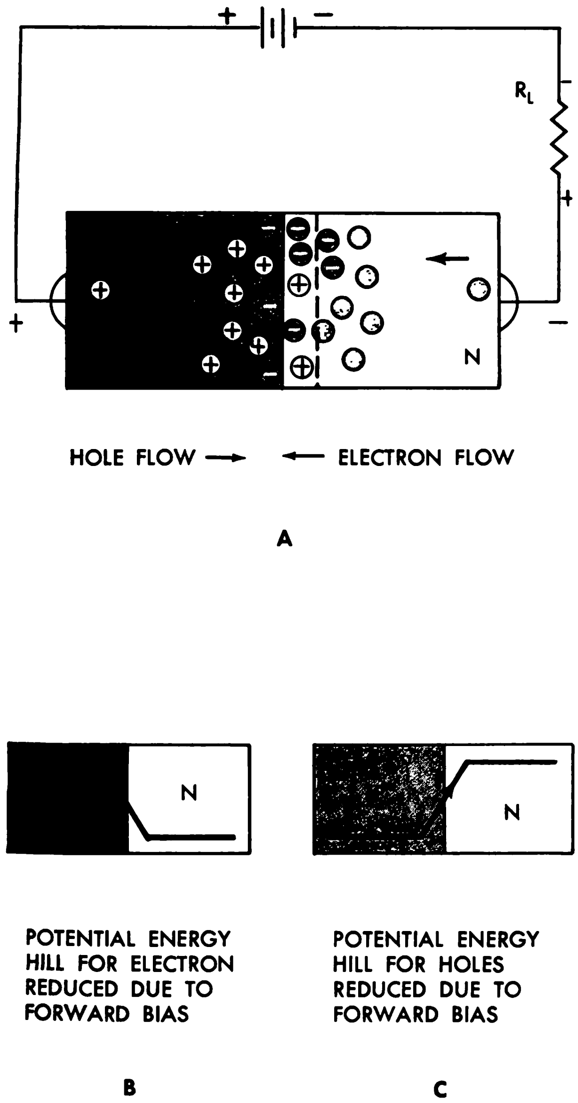
Holes and elections flow into their respective regions of lower potential energy more readily and meet greater opposition when trying to climb their high potential energy hills. Hence, current flow in N-type germanium is mainly a flow of electrons, while that in P-type germanium is mainly a flow of holes. The amount of current flow that can take place through a P-N junction depends on the polarity as well as the magnitude of applied bias voltage.

P-N JUNCTION WITH FORWARD BIAS. If the negative terminal of applied voltage is connected to the N-type germanium as shown in the next illustration at A, the negative terminal of the battery repels the electrons and causes them to migrate toward the P-type germanium. In a similar manner, the positive terminal of the battery repels the holes and causes them to migrate away from the positive terminal toward the junction.

As shown in the accompanying figures B and C, the potential energy hills for electrons and holes has been reduced due to forward bias. Hence, some electrons enter the P-region and some holes enter the N-region.

The junction, under these conditions, permits a free exchange of positive and negative charges (holes and electrons), decreasing the impedance of the junction and causing a comparatively large current flow to take place from the negative terminal to the positive terminal in the direction indicated.

In other words, electrons and holes combine in a small area of diffusion on either side of the P-N junction between the dotted lines in the illustration. For each hole in the P-region that combines with an electron



P-N Junction with Forward Bias

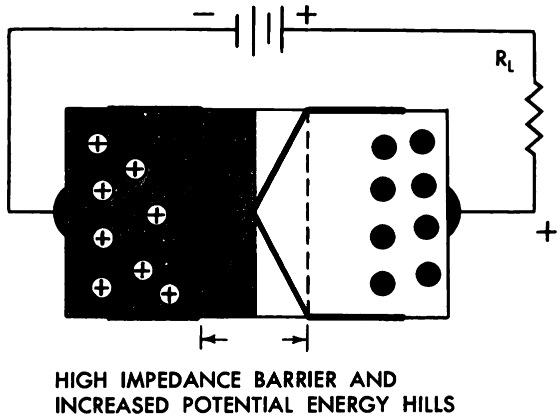
from the N-region, an electron from an electron-pair bond in the crystal near the positive terminal of the battery enters the battery at the positive terminal.

This action creates a new hole in the P-crystal which, in turn, migrates toward the N-type germanium. In a similar manner, for each electron that combines with a hole in the N-type germanium, an electron enters the N-crystal from the negative terminal of the battery. The current flow in the P-crystal is mainly a flow of holes, while that in the N-crystal is mainly a flow of electrons.

We can say then that, under these conditions, the germanium diode is biased in the low impedance or forward direction and that the junction of the P- and N-type germanium wafers acts like a good conductor.

Let us now consider the action when the DC supply voltage is reversed.

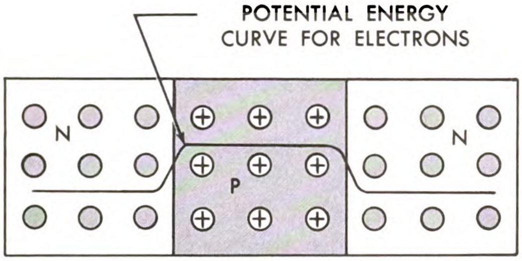
P-N-JUNCTION WITH REVERSE BIAS. If the positive terminal of the voltage source is connected to the N-type germanium, the free electrons will be attracted toward this terminal and away from the junction as shown in the accompanying illustration. In a similar manner, the negative terminal of the battery attracts the holes and causes them to migrate away from the junction toward the negative terminal of the power source. Thus, with the DC source voltage reversed, electrons and holes concentrate near the battery terminals with very few present at the junction. Under these conditions an exchange of positive and negative charges is quite difficult, and very little current flow can take place past the junction. In other words, the potential energy hill for holes and electrons has increased due to reverse bias. The junction offers a higher barrier impedance to current flow, and the semiconductor acts somewhat like an insulator with practically all of the potential drop across the high impedance barrier at the junction.



P-N Junction with Reverse Bias

Germanium Triodes

Germanium triodes of the junction type usually consist of three wafers of semiconductor material with the center wafer of opposite type material from the two outer wafers. If the outer wafers are P-type germanium and the center wafer is N-type germanium, the unit is called a P-N-P junction transistor. And, as the letters indicate, an N-P-N transistor consists of a thin wafer of P-type germanium between two wafers of N-type germanium. The operation of P-N-P transistors is similar to that of the N-P-N type except that the hole constitutes the main current component instead of the electron. Hence, only the N-P-N types will be discussed in detail

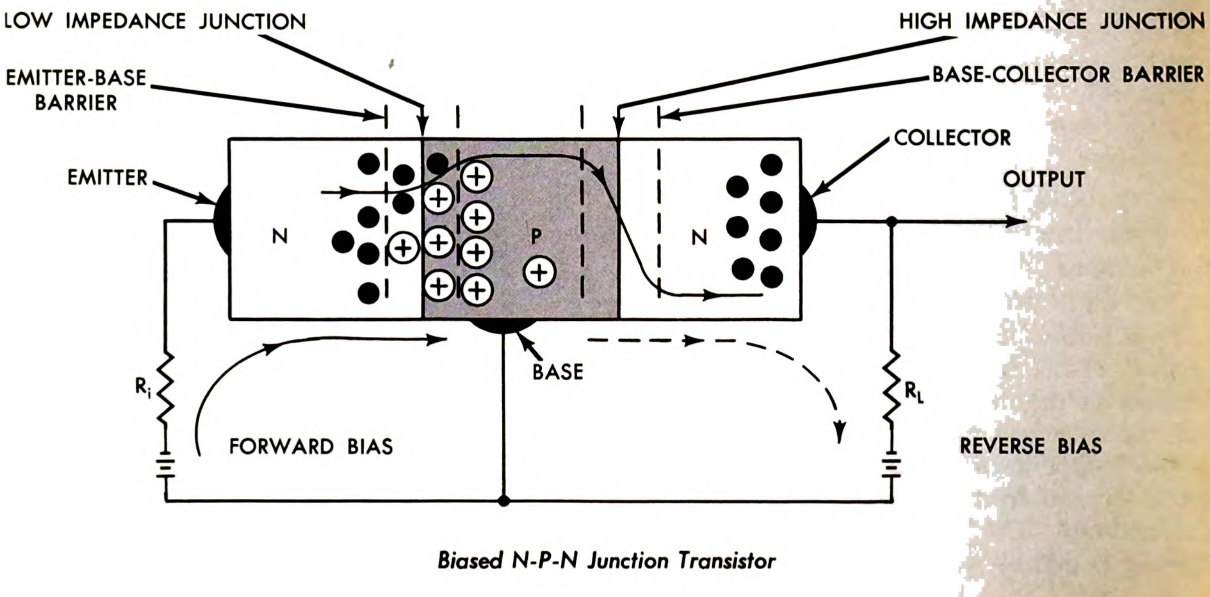


N-P-N Junction Transistor in State of Equilibrium

N-P-N JUNCTION TRANSISTORS. The illustration shows the distribution of holes and electrons in an N-P-N junction transistor under equilibrium conditions with no external voltages applied. Note that electrons are concentrated in the region of lowest potential energy for them and cannot climb the potential energy hills to enter the P-type germanium. The holes are concentrated in their lowest potential energy region also. However, electrons are the major current components in N-P-N transistors, so potential energy diagrams for holes are not shown.

BIASED N-P-N JUNCTION. Let us now assume bias voltages are connected to an N-P-N junction transistor as illustrated in the accompanying diagram. In the circuit arrangement shown, the left-hand junction between the emitter and the base is biased in the forward or low impedance direction. The right-hand P-N junction between the base and the collector is biased in the reverse or high impedance direction.

There is very little potential drop across the emitter-base barrier, practically no potential difference exists in the base region, and practically all of the potential drop between the external base and collector leads occurs across the base-collector barrier.



Biased N-P-N Junction Transistor

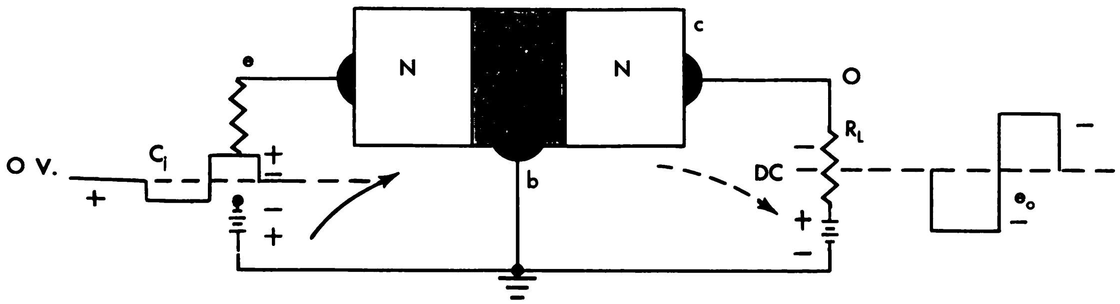
The negative bias current causes electrons to migrate from the emitter toward the base, reducing the potential energy hill at the left-hand junction thus allowing some electrons to climb this hill and enter the P-type germanium. In order for emitter current to cause a change in collector current, electrons that have migrated from the emitter to the base must traverse the base region and appear at the base-collector barrier. Since there is virtually no electronic difference in potential within the base region, the process of electron travel is essentially one of diffusion owing to the greater concentration of holes and electrons in the base region near the emitter than in the base region near the collector. In other words, the base wafer is relatively thin, and most of the electrons which enter it will not combine with the holes; instead they pass through the wafer and are readily attracted down the potential-energy slope shown at the right in the referenced illustration. The steep potential-energy slope which permits easy entrance of electrons from the base into the N-type germanium of the collector is produced by the application of reverse bias between collector and base.

Consequently, electrons flow freely from the emitter to the base, limited only by the series resistor R_i, and produce or inject an excess of electrons into the base wafer as shown by the solid arrows. Current flow takes place through the right-hand junction because of the excess electrons injected into the base through the left-hand junction. These electrons injected into the base virtually reduce the high impedance of the right-hand junction, allowing current flow to take place as indicated by dashed arrows in the illustration.

Although the magnitude of DC current through the right-hand junction may not be any greater than that flowing through the left-hand junction, the output DC voltage across the load resistor R_L may be considerably greater than the DC voltage across the input resistor R_i. This is true because of output resistor R_L being considerably larger than R_i.

N-P-N JUNCTION WITH SIGNAL APPLIED. If current flow through the left-hand junction is varied, the number of excess electrons injected into the center material and the impedance of the right-hand junction will vary. This, of course, results in a similar current variation through the right-hand reversed bias junction and a proportional change in the output voltage across R_L.

For example, consider the diagram, *A Grounded Base N-P-N Junction Transistor,* where the current flow through the left-hand junction is varied by adding a small AC signal in series with the DC bias voltage. With the negative portion of the input signal applied, the total emitter-base voltage becomes larger in the forward direction, making the emitter more negative with respect to base. This, of course, augments emitter-base current flow, which increases the number of electrons injected into the base. This decreases the impedance of the right-hand junction (base-collector barrier), permitting



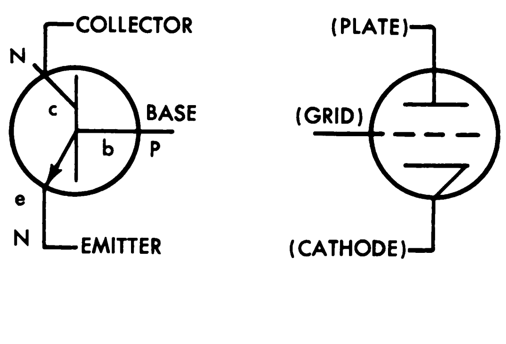
Grounded Base N-P-N Transistor with Signal Applied

a greater base-collector current flow which, in turn, results in a greater negative drop across the output load resistor.

When the positive portion of the input signal is applied, it opposes the forward bias and makes the emitter less negative with respect to the base. Hence, it increases the barrier impedance of the emitter-base and, in turn, the base-collector junction, resulting in a smaller base-collector current and a smaller negative drop across R_L.

Note that the illustration shows an in-phase amplified signal appearing across the collector load. It must be remembered, however, that it is the difference between the input and the output impedances that permits gain to be obtained in a grounded base transistor amplifier, since the emitter and collector currents may be very nearly the same. The actual current amplification in a grounded base circuit is always less than one.

N-P-N (P-BASE) POINT-CONTACT TRANSISTOR. There is a second method of transistor construction, resulting in a unit called a *point-contact* transistor. As shown in the accompanying illustration, the internal connections to the emitter and collector are made through the ends of very small wires in contact with the surface of the crystal at points very close together. The high-resistance, small-area contacts on the crystal surface are separated by only a few thousands of an inch.

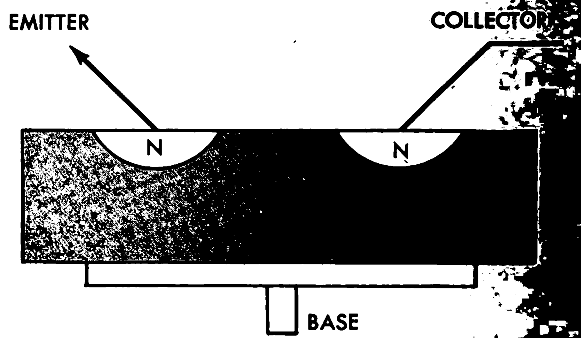


N-P-N or P-Type Point-Contact Transistor

The figure illustrates a P-type crystal (main body) for which there is a forming process during manufacture that produces at each contact position a small area of N-type material. Hence, the P-type point-contact transistor is similar in operation to the N-P-N junction type. However, point contact units will provide faster switching action than junction triodes and are more desirable for high-speed computer switching circuitry. Point-contact transistors are generally utilized in the grounded base amplifier circuits of the type discussed on page 4-71.

Comparison Between Transistor and Vacuum Tube Operation

The conventional *N-P-N Transistor Symbol and Its Likeness to the Vacuum Triode* is illustrated. The base is represented by a



straight line with the emitter and collector represented by lines at angles to the base. Note the slanting line for the emitter is marked with a small arrowhead which points *away from* the base in the direction of *hole flow* (electrons flow against the arrow).

Practically all the electrons that emerge from the emitter (cathode) flow to the collector (plate). The base (grid) current is minute and consists only of a small number of electrons which represent the recombination number of holes and electrons in the P region.

Similar circuit arrangement may be used with both transistor and vacuum tube amplifiers. However, since the transistor is

cally a current operated unit, it operates as a current amplifier and prefers a constant current power supply. The vacuum tube, on the other hand, operates as a voltage amplifier and prefers a constant voltage power supply.

As you well know, most vacuum tube circuits feature both high input and output impedances. On the other hand, transistor circuits usually have considerably lower input impedance with respect to their output impedance. However, their input and output impedance depends on the manner in which they are utilized in various circuits.

As with vacuum tube amplifiers, there are three basic transistor amplifier circuits, each of which is identified by the grounded or common element. That is, transistor amplifiers are identified as either a grounded base (grid), grounded emitter (cathode), or grounded collector (plate) amplifier.

The grounded base amplifier was discussed on page 4-71.

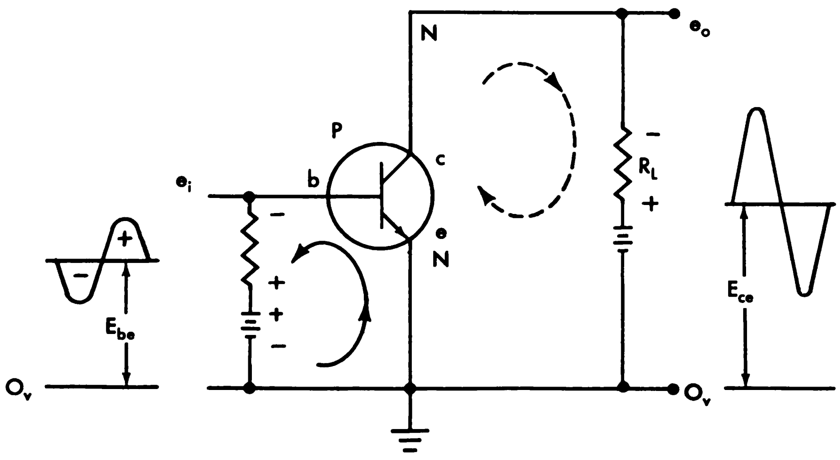
GROUNDED EMITTER AMPLIFIER. A very common transistor circuit known as the grounded emitter amplifier is shown in simplified schematic form.

In this amplifier, the average base-emitter potential is determined by the forward bias voltage E_{be}, which allows some average emitter-base current flow (dashed arrows). This emitter-base current flow injects free electrons into the base, decreases the impedance of the base-collector junction, and allows an emitter-collector current to flow (solid arrow).

A small direct current applied to the base-emitter circuit can control a much larger current in the emitter-collector circuit. In practice, current amplification of ten or twelve to one is easy to obtain with a grounded emitter transistor. An emitter-base current of 0.1 ma. may result in a current through the load resistor of 1.0 ma. or more.

When a signal is applied to the input, the bias current varies about its average value, and corresponding variations, though of much greater magnitude, take place in the collector circuit. This results in an amplified version of the input signal appearing across the load impedance or between the collector and the common base.

Note that a phase reversal occurs between the input and output of the grounded emitter circuit. That is, when the negative-going portion of the signal pulse is applied at the input so that it opposes the forward bias as shown, the emitter-base current will decrease. This results in few electrons injected into the base which increases the impedance of the collector circuit. This, in turn, results



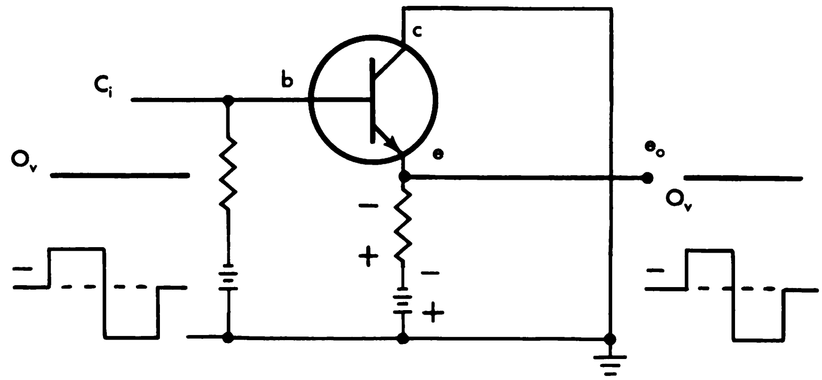
Grounded Base P-N-P Junction or P-Base Point Contact Transistor

in a much smaller emitter-collector current, which causes a much smaller drop across the load resistance (less negative) and allows the voltage between collector and ground to become much larger (more positive).

GROUNDED COLLECTOR AMPLIFIER. Transistorized versions of the vacuum tube cathode follower utilize the common collector type circuit. One variation of a common collector circuit is shown in the next illustration.

The common collector transistor circuit, like the vacuum tube cathode follower, has high input impedance and low output impedance and may, therefore, be used as an impedance matching device. Its current amplification can be made rather large, but its voltage gain is always less than one.

BIAS VOLTAGES. The proper bias voltages depend on the type of transistor used and which terminal is to be grounded or used as



Common Collector Amplifier

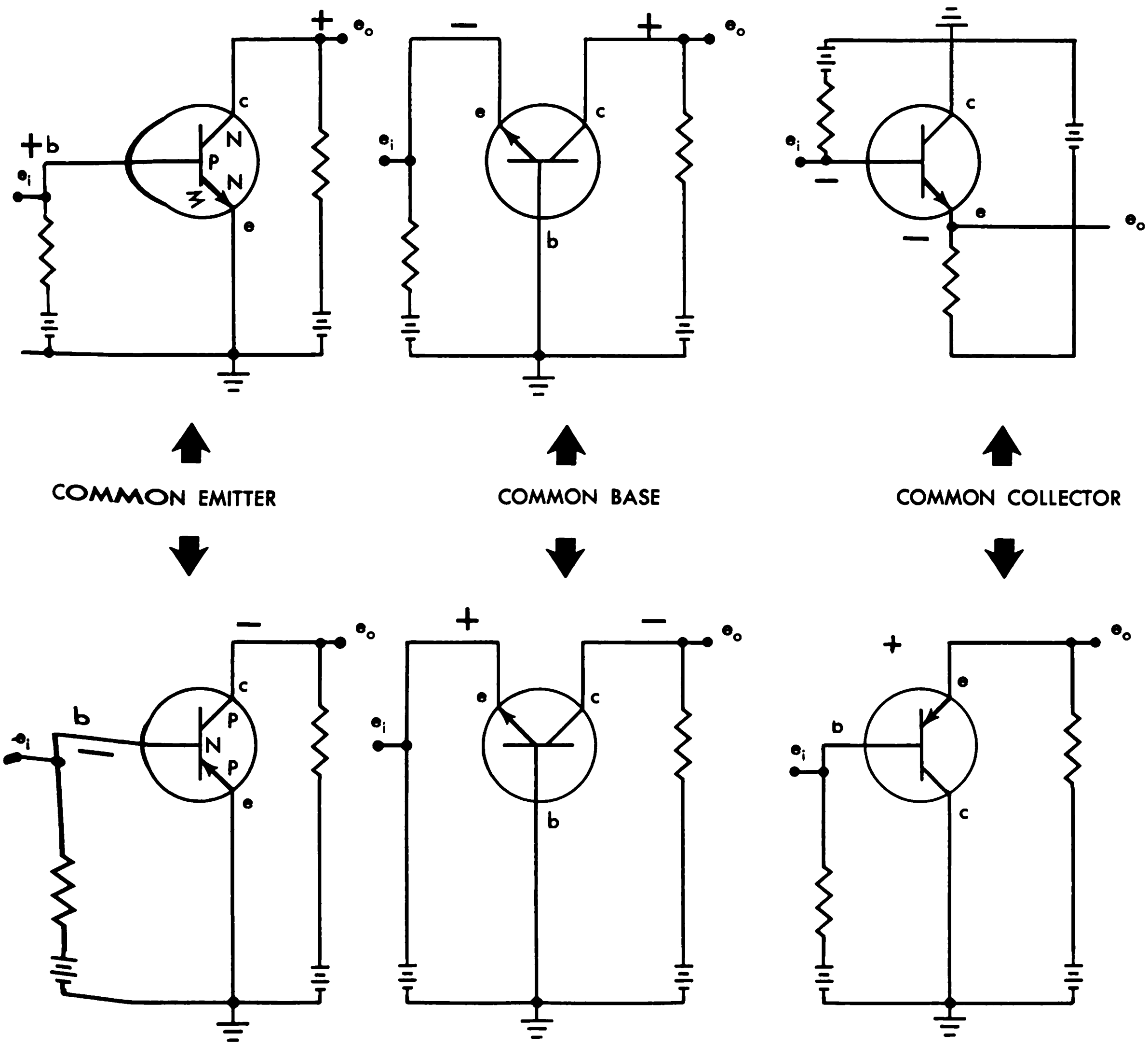
When the positive-going portion of the input signal is applied between base and ground, it aids the forward emitter-base bias, augments emitter-base current flow, and increases the number of electrons injected into the base. This, in turn, decreases the impedance of the base-collector barrier, allowing a greater emitter-base-collector current flow through R_e, which produces an in-phase positive-going output pulse between emitter and the grounded collector.

Upon careful investigation, you will note that the positive-going pulse across R_e decreases the negative potential at the emitter with respect to the base, thus opposing the forward emitter-base bias and degenerating the input signal. You will recall that this degeneration is similar to the negative feedback resulting in a conventional vacuum tube cathode follower because of the unbypassed cathode resistor.

reference. The diagrams illustrate the bias polarities required for the three basic types of amplifier circuits and for different transistor types.

As you investigate these diagrams, keep in mind the fact that the emitter is biased in a forward direction with respect to the base in all examples. All polarities indicated are with respect to the grounded or common element.

The upper row of diagrams illustrates the required bias polarities for N-P-N junction of P-base point-contact type transistors. Note that the collector-emitter polarities are similar to the plate-to-cathode polarities encountered in vacuum tube amplifier circuits. That is, just as the plate of a triode tube is made positive with respect to its cathode, so is the collector of an N-P-N transistor made positive with respect to its emitter.

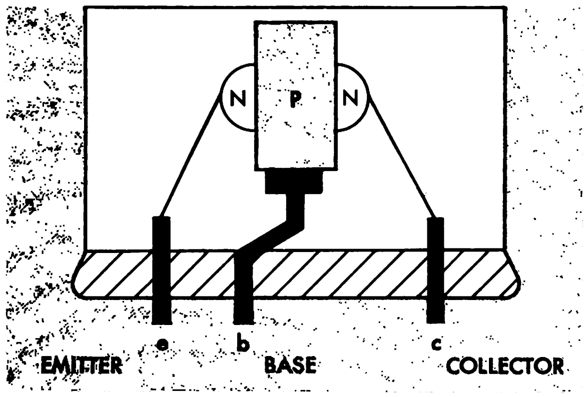


The lower row of diagrams shows the required bias polarities for P-N-P junction or N-base point-contact transistor types. It is extremely important to remember that the major conduction in P-N-P type transistors is by holes instead of electrons; hence, their bias polarities are exactly opposite to those shown in the top row of diagrams. Consequently, the proper bias polarities must *always* be observed when working with transistorized circuits. Note that the arrow head in the symbol for the P-N-P type transistors is also reversed—pointing *toward* the base—but still in the direction of hole flow.

TERMINAL LEAD CONNECTIONS. Special care should be exercised to check transistor types whenever repairing transistorized equipment or assembling new circuits because improper application of bias supply voltages—even if only momentarily—may destroy a transistor.

The accompanying diagrammatic sketch showing structural arrangement of a typical N-P-N junction triode also illustrates the presently accepted standards for terminal lead connections. The pin size and unequal spacing is chosen in order to permit the use of standard subminiature tube sockets. Some

transistors are supplied with long leads to permit the units to be wired directly into the circuitry where used. If socket mounting is desired, the leads may be cut shorter.



Structural Arrangement of Terminal Leads

The illustration of structural arrangement shown is not to scale. All commercially available transistors are physically quite small with overall dimensions, exclusive of terminal leads, seldom exceeding one-half inch.

As we have seen, transistors may be divided into two general types—point-contact and junction—depending on the method of construction employed. Both basic types may be further subdivided. Point-contact transistors can be classed as having either a P-type or on N-type base. Junction transistors may be classified as N-P-N or P-N-P types.

In P-N-P junction or N-type base point-contact transistors, the power supply voltage polarities are reversed with respect to the N-P-N types. Also, you will remember their main conduction is by holes instead of electrons. Otherwise, the operation of the P-N-P transistor types is similar to that of the N-P-N types discussed in more detail.

Generally speaking, junction transistors are capable of giving more power gain than point-contact types when both types are used in similar circuits. However, gain does depend on the basic circuit employed. The grounded emitter and grounded base circuits are used where signal amplification is desirable. The grounded emitter circuit is the more popular when junction transistors are used, and grounded base the more désirable when point-contact transistors are used. The grounded collector circuit is primarily used for an impedance matching device, the transistor substitute for the cathode follower.

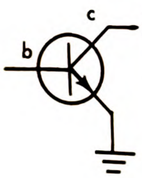
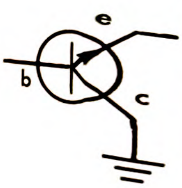
The chart summarizes the general characteristics of the three basic types of triode transistors.

SATURABLE REACTORS

We know that perfect operation of important airborne equipment can be hindered by the sudden failure of a vacuum tube. Shocks and vibrations damage tubes, and their life is very limited in practical airborne equipment. These serious disadvantages of vacuum tube amplifiers were considered, and a tubeless amplifier was developed in addition to the transistorized amplifiers.

Amplifiers using transistors are based upon the possibility of controlling the resistances of thin semiconductor wafers by the application of electric fields strong enough to penetrate their surfaces. When two point contacts are placed close together on the surface of a small block of germanium, the output voltage and current from the collector contacts can be controlled by the passage of current through the emitter control contact. That is, by changing the control emitter current, we alter the resistance of the output collector circuit. Probably for this reason the term TRANSISTOR was coined from TRANSient-resISTOR. In a somewhat similar manner, by changing the control current of a saturable reactor we alter the inductance of its output circuit. Hence, the term TRANSDUCTOR was coined from TRANSient-inDUCTOR.

Since their inception, transductors—saturable core reactor devices—have proved to be

	INPUT IMPEDANCE	OUTPUT IMPEDANCE	POWER GAIN	CURRENT AMPLIFI-CATION	FREQUENCY RESPONSE	PHASE RELATION-SHIP
 COMMON EMITTER	MODERATE	MODERATE TO HIGH	HIGH	LARGE	MODERATE	180 PHASE REVERSAL OCCURS BETWEEN INPUT AND OUTPUT
 COMMON BASE	VERY LOW	HIGH	MODERATE TO HIGH	LESS THAN 1.0	POOR	IN PHASE WITH INPUT
 COMMON COLLECTOR	HIGH, BUT DEPENDS ON LOAD	MODERATE TO LOW	LOW	LARGE	GOOD	IN PHASE WITH INPUT

Summarizing the Characteristics of the Basic Types of Triode Transistors

very useful components for numerous applications in airborne electronic equipment. They have been developed into extremely valuable arrangements which are applied in telemetering and, in connection with transistors, in servo control applications. Among the most interesting applications in this latter category are those in which transistors are used as voltage amplifiers with magnetic amplifiers as power stages.

The saturable reactor is a natural complement to the transistor since it also lends itself readily to hermetic sealing, miniaturization, and rugged construction. Semiconductor devices also permit matching low-impedance magnetic amplifier outputs to high-impedance servo controls.

The combined characteristics of transductors, transistors, and other solid state electronic devices are presently affording many opportunities for new control devices. Because of the economy and reliability of these solid state devices, more routine manual operations in weaponry are becoming automatic. Hence, solid state elements are rapidly becoming common components in contemporary interceptor aircraft and other airborne equipment.

Principles of Operation

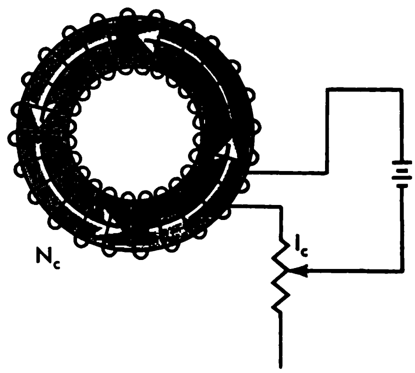
The operation and performance of saturable reactors is largely dependent on a few basic rules of magnetism, the properties of special magnetic core materials, and the manner in which these cores are wound. The necessary rules of magnetism, the properties and characteristics of special magnetic core materials, and Ohm's law for magnetic circuits are assumed to be understood; however, a brief review of the B-H curve and permeability of ferromagnetic materials will be discussed prior to examining a transductor-saturable reactor.

The region in the vicinity of a magnet where it exerts force is called a *magnetic field*. The magnetic field strength (H) at any point in space is defined as the force that the field exerts upon a north pole of unit strength located at that point. The magnetic field strength is also called the *magnetizing force (mmf)*. It is considered as the force which tends to line up all of the molecular magnets in the direction of the magnetizing force when a magnetic material is placed in a magnetic field. The opposition to the alignment of these molecular magnets is due to molecular friction and is called *reluctance*. The total number of lines of flux inside the magnetic material per unit area is called *flux density* or *magnetic induction (B)*. As the magnetizing force H is gradually increased from zero to large values, it overcomes the reluctance of the material and allows the flux density to increase. When the magnetizing force overcomes the reluctance in this manner, it is said to be decreasing the reluctivity of the material or increasing the permeability of the material.

$$permeability = \frac{1}{reluctivity}.$$

The permeability or reluctivity of a material cannot be stated even if its composition and heat treatment are known, because flux does not change in direct proportion to magnetizing force or ampere-turns.

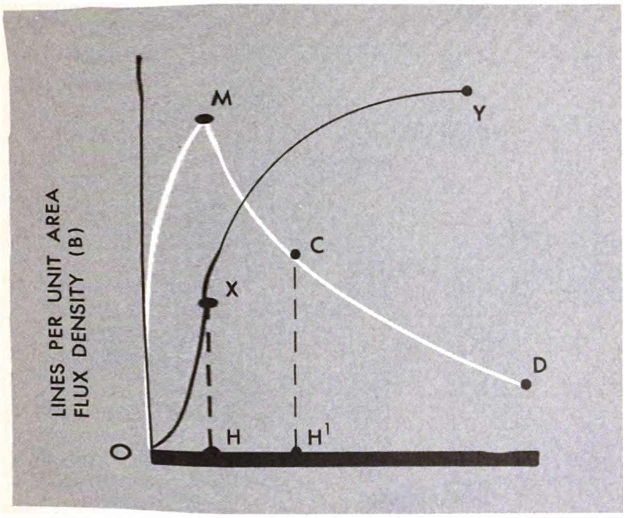
To make this clear, suppose we consider a steel ring surrounded by a coil as illustrated in the *Basic Magnetic Circuit for Controlling Magnetizing Force.*



Basic Magnetic Circuit for Controlling a Magnetizing Force

As the rheostat is moved upward, the DC control current I_c flowing through the coil N_c increases. This increases the magnetizing force which, in turn increases the lines of magnetic induction—this is indicated by the *magnetization curve* in the accompanying illustration.

It should be noted that the flux increases slowly at first, but soon rises rapidly. This is due to the higher magnetizing force over-



Magnetization Curve

coming the reluctance of the core and increasing its permeability. It should be further noted that the flux density rises most rapidly at about point X, at which time the permeability is approaching its maximum value—at point M in the permeability curve. It is important to remember that it is within this region where the permeability is maximum that the magnetizing force produces the largest inductance for the coil.

As the magnetizing force increases from point H, the flux density continues to increase beyond point X, but more gradually as it approaches saturation. When the magnetic induction in a piece of magnetic material approaches its maximum value, as at point Y, the metal is said to become saturated with magnetic lines of flux. That is, there is little space available for more flux lines, and the core presents a greater reluctance to an increase in the number of lines of flux.

The permeability curve shows that, as the magnetizing force is increased beyond points H to H^1, the permeability decreases rapidly from M to C. This shows that an increase in magnetizing force beyond point H becomes less effective in producing flux and thereby results in a smaller inductance for the coil as saturation is approached. It is this characteristic of variable inductance of a coil, when its core is subjected to various degrees of magnetization, that makes the saturable reactor such a valuable device.

Basic Winding Arrangements

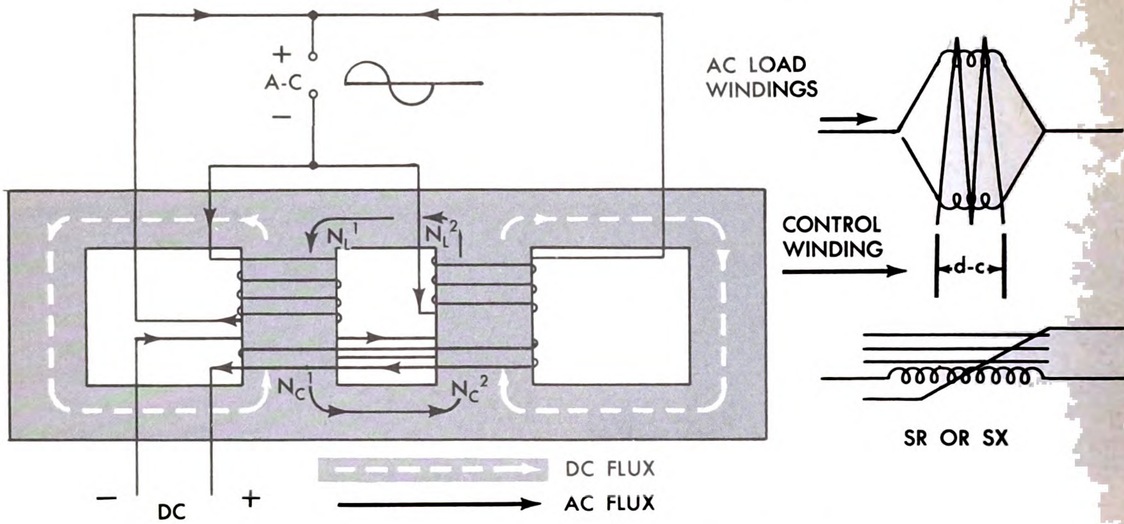
The saturable reactor is a variable inductance that looks like a transformer with two or more separate windings on an iron core. The figure illustrates the *Arrangement of Windings on a Saturable Reactor and its Common Schematic Symbols.*

The performance of magnetically controlled saturable reactor devices depends largely on the arrangements of the various windings. These windings serve as load windings, control windings, feedback windings, or as additional windings for special purposes.

It is generally desirable that all windings be placed on their cores in such a way that leakage effects are reduced to a minimum. Another important requirement is to prevent the circulation of disturbing currents of the AC winding from affecting the DC control, bias, or feedback circuits. In the arrangement of windings illustrated, half of the AC windings are wound on each of the two inside legs. These half windings, $N_L{}^1$ and $N_L{}^2$ are connected in parallel and wound in such a way that when the AC current assists the magnetizing force in one core, it opposes the DC magnetizing force in the other. Hence, AC voltages induced into the DC control windings $N_C{}^1$ and $N_C{}^2$ are series-opposing and are cancelled out; therefore, there is no resultant AC in the DC bias or control circuit.

A THREE-LEGGED CORE. A three-legged core saturable reactor element whose center leg is split is illustrated next. Note that this core arrangement has only one DC control winding which embraces both magnetic circuits. There is no resultant AC flux linking the DC winding, and no net voltage can appear across the control circuit terminals.

In this arrangement there are two AC coils ($N_L{}^1$ and $N_L{}^2$) but now wound in series

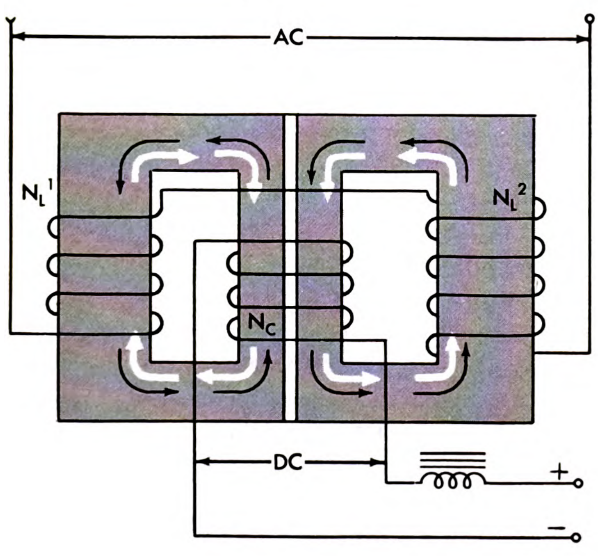


Arrangement of Windings on a Saturable Reactor and Its Common Schematic Symbols

on the outer two legs of the common core. Again the AC flux of one coil opposes the DC flux, while the AC flux of the other coil aids the DC flux on alternate half-cycles.

If the center leg were not split, the AC flux components produced by the AC windings would tend to flow around the circumference of the core and not through the center leg at all. Hence only unidirectional flux would flow through the center leg and cause disturbing hysteresis effects.

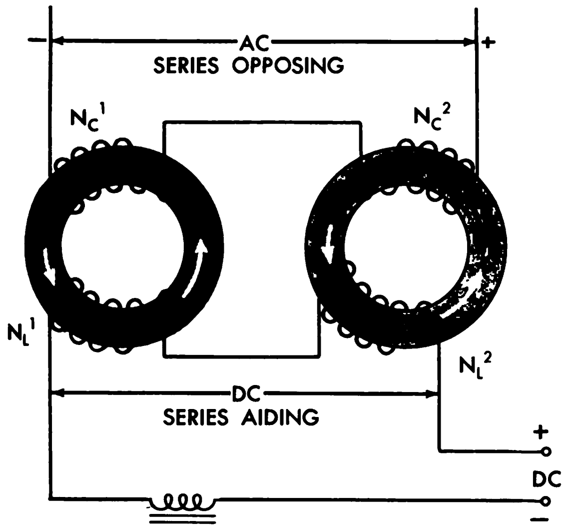
By splitting the leg lengthwise, a narrow air gap is provided between the cores. This prevents the AC flux from following the circumference of the core and permits it to flow through both halves of the center leg —always in opposite directions to prevent inducing AC voltage in the DC control winding.



A Three-Legged Core

RING-TYPE REACTOR ELEMENTS. Another arrangement of core and coil construction providing two separate *Ring-Type Saturable Reactor Elements* is shown in the accompanying diagram. In this diagram, both DC coils are series-aiding, and the AC coils are series-opposing. With this arrangement, the two AC voltages which are induced into the control windings N_C^1 and N_C^2 cancel each other out, and there is no resultant AC current in the control circuit. However, it is important to note that the actual voltage induced into each individual DC winding could be very large if the turns ratio of N_C/N_L were large. This would require a high

degree of insulation and thereby require a comparatively large space for the insulating material.



Ring-Type Saturable Reactor Elements

In order to impede possible induced AC currents in the control winding, an inductance choke may be added in series with the DC circuit as shown in the previous illustration.

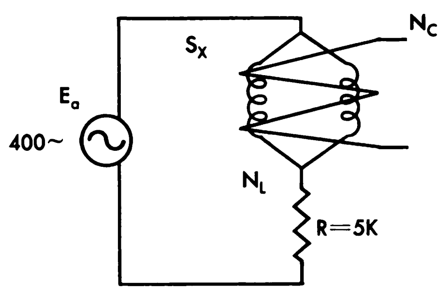
Another method for eliminating undesired effects of transformer action between AC load and DC control windings consists of winding the AC coil at an angle of 90 degrees to the plane of the control coil so the mutual inductance between the two coils is practically zero.

Effect of DC Control Current. When no DC control current flows through the direct current windings, a reactor has a large inductance and acts as a choke limiting the flow of alternating current through its AC windings. However, when even a few milliamperes of direct current flow through the many turns of the DC winding, the iron core becomes saturated and loses its inductive effect. It is then less able to prevent AC from flowing through its load windings. Stated differently, we increase the flow of alternating current through the saturable reactor when we increase the amount of current in its DC control winding.

When a saturable reactor is used in control circuits such as for phase shifting (explained below), it may be smaller than a man's fist. When it is used directly to increase or decrease the AC voltage applied to heavy loads, it may be rated from one to several hundred KVA and may weigh hundreds of pounds.

Basic Reactor Control Circuit

Let us consider the *Basic Reactor Control Circuit* illustrated, consisting of the AC windings (N_L) of the saturable reactor (S_X) connected in series with the resistor across a steady AC voltage. When no direct current flows through the DC control winding, this reactor has a large inductance or choke effect which prevents the flow of much alternating current.

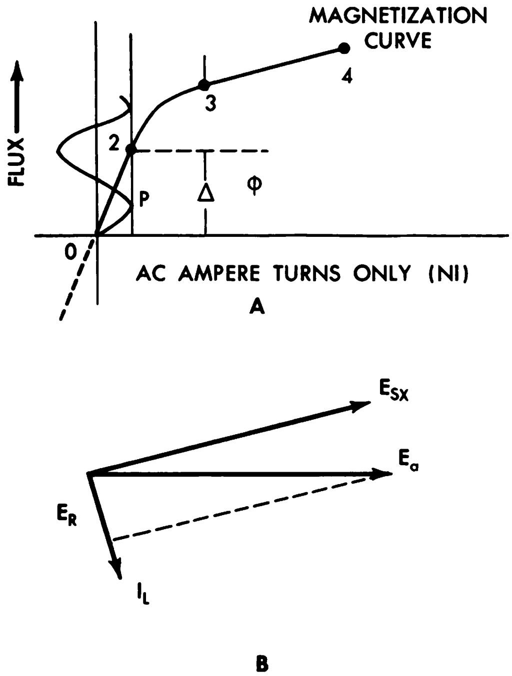


Basic Reactor Control Circuit

With the 400-cycle AC applied, let us assume the action of the reactor is such that it permits the same amount of current to pass through it as would flow through a 30,000-ohm inductive reactance. When about 2 ma. of direct current flows through the many turns of the control winding, the core

of the reactor becomes saturated to the extent that the inductance decreases so that it passes as much 400-cycle current as a 1,000-ohm reactance.

OPERATION BELOW THE KNEE OF SATURATION. The first mentioned action of a saturable reactor—when no direct current flows in the DC winding of S_X—is illustrated in the accompanying diagrams A and B.



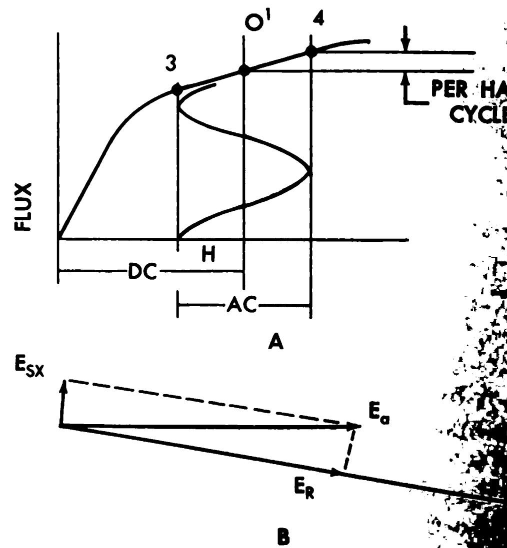
Saturable Reactor Action-Operating Below the Knee of Saturation

The magnetization curve at A shows how the flux in the reactor core of S_X increases as the current in either winding increases. With no DC flowing through the control winding, the only magnetic flux present is caused by the small current flowing in the AC load windings. The sine wave shown superimposed on the magnetization curve represents the magnetizing force (NI) produced by the AC reactor current and the number of AC coil turns.

When this magnetizing force is zero, as at point *0*, there is no flux. Once each cycle, the AC current increases the magnetizing force (ampere turns) to its peak as indicated at point *p*. This increases the flux to point *2* on the magnetization curve. It is this large change in flux (Δ Φ) from *0* to *2* each half-cycle that causes the choke effect which limits the amount of AC current in S_X. Under this condition S_X has a large inductance and, hence, a large reactance to AC.

The large ratio of inductive reactance to resistance in this basic circuit causes the AC current *I* to lag the applied voltage by approximately 90 degrees as illustrated by the accompanying vectors at B in the illustration. Note also that the AC voltage across S_X is much greater than that across *R*, and that it is very nearly in-phase with the applied voltage.

OPERATION ABOVE THE KNEE OF SATURATION. Suppose the same saturable reactor is initially energized with a direct current (as shown at A in the next illustration) from *0* to *0′* well above the saturation bend on the magnetization curve. The point *0′* represents the amount of flux (*HO′*) produced by the magnetizing force or current (*OH*).



Saturable Reactor-Operating Above the Knee of Saturation

If an AC applied voltage of the same magnitude as before is used to supply the AC load component of current through S_X, the current in the AC winding will produce a sinusoidal variation of magnetizing force as shown. Notice that the AC component of magnetizing force still remains well above the saturation knee points *3* to *4* on the magnetization curve. Notice also that the peak change of magnetizing force during each half-cycle produces only a very small change in magnetic flux. This flux cannot change more, for here within region *3* to *4* the iron core is nearly saturated or almost completely magnetized.

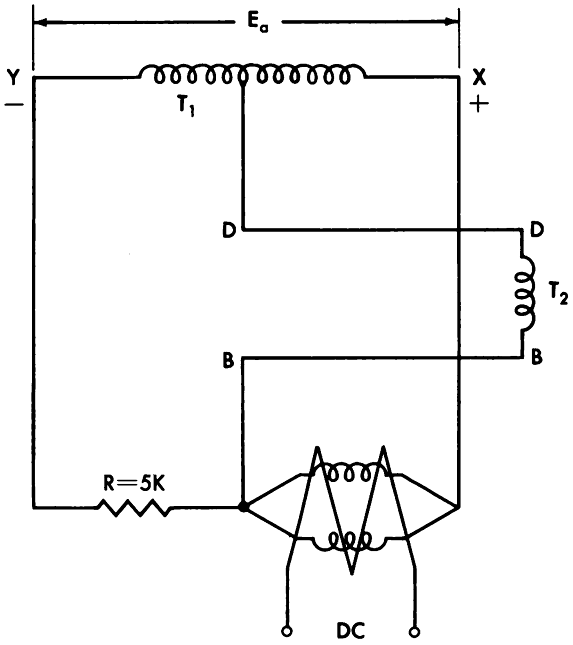
Since the flux does not change much in the region above the saturation knee, the inductance of the saturated reactor is much smaller. It has a smaller reactance and very little choking effect. Hence, much more alternating current may flow through the AC winding for a given applied voltage.

The reactance of S_X is now small with respect to the series resistance, and the greater current produces a much larger voltage across this series resistance. These conditions are illustrated by the vectors at B in the diagram. Notice, in this case, that the load current is now nearly in-phase with the applied voltage and that the voltage produced across the AC coils of the reactor is leading the applied voltage by nearly 90 degrees.

Saturable Reactor Phase Shifting Bridge

Now consider modifying the basic reactor control circuit, previously described, to the one illustrated representing a *Saturable Reactor Phase Shifting Bridge.* Note that the only change in the circuit is the secondary of a transformer T_1 acting as the supply voltage and that a load T_2 is placed between points *DB.*

The vector diagram showing the phase relationship between reactor current, reactor voltage, and applied voltage—for the condition of no DC control current—was illustrated on page 4-82. The vectors are

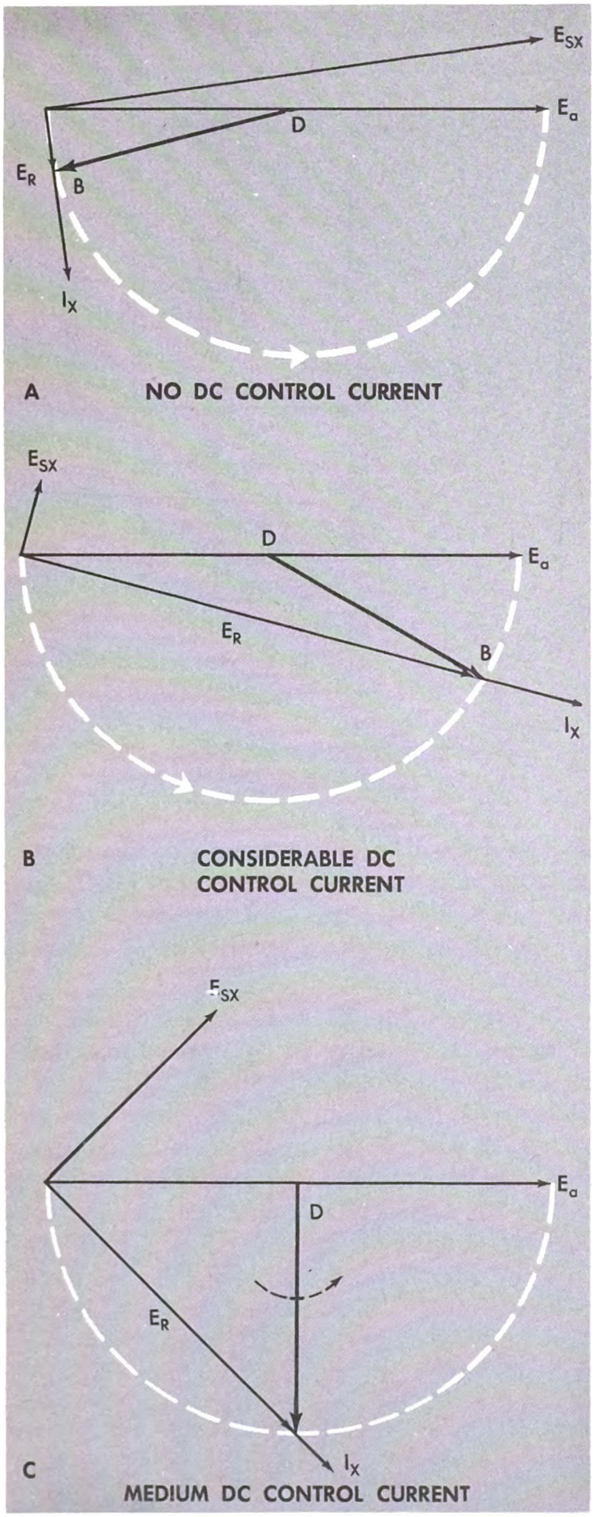


Saturable Reactor Phase Shifting Bridge

redrawn at A in the sequence of diagrams showing also the *Phase Relations Between the Load Voltage DB and Applied Voltage* for the three indicated conditions of control current.

CONDITION OF ZERO CONTROL CURRENT. With no DC flowing in S_X, the voltage *DB* is a sine wave nearly 180 degrees out-of-phase with the AC supply voltage as though points *B* and *Y* were connected together, since the resistance is negligible with respect to the reactance of S_X. With no DC control current flowing in the control winding, the core is unsaturated, representing a large effective iron, a large inductance, and hence a large reactance to AC. This prevents the flow of much alternating current through the AC windings, producing a small voltage across *R* (represented by the short voltage arrow), and a comparatively larger voltage across the reactor, E_{SX}.

The voltage across the load is, of course, equal to the vector sum of one-half of the



Phase Relations Between Load Voltage DB and Applied Voltage

applied voltage, and the voltage across the resistor as shown in the vector diagram.

CONDITIONS OF CONTROL CURRENT FLOWING. When DC is applied to the control winding, it decreases the effectiveness of the iron core, decreases the inductance of the AC coil, and makes the reactor less able to prevent the flow of alternating current through its AC windings.

Since the inductive reactance of S_X is decreased and R remains constant, the vector voltage across S_X will decrease and that across R will increase as illustrated at **B** and **C** in the accompanying vector diagrams. As a result of these changes, the vector DB (voltage across T_2) swings counterclockwise along its circular locus. Note how it now lags the applied voltage by much less than before.

With a Large D.C. Control Current Saturating S_X, the load voltage DB will be shifted nearly 180 degrees as illustrated at **B.** Notice also that, in such a case, the load voltage DB is nearly in phase with the applied voltage as though points X and B were connected together because the reactance of S_X is negligible with respect to the resistance R.

If a Medium Amount of D.C. Control Current—between the two extreme conditions just discussed—were to flow through the DC winding, causing the reactance of the AC winding to be equal to the series 5,000-ohm resistor, the AC voltage across S_X would be equal to that across R. That is, $E_{SX} = E_R$. This condition is illustrated at **C** in the series of graphical representations. Note that at C the voltage DB lags the applied voltage by 90 degrees.

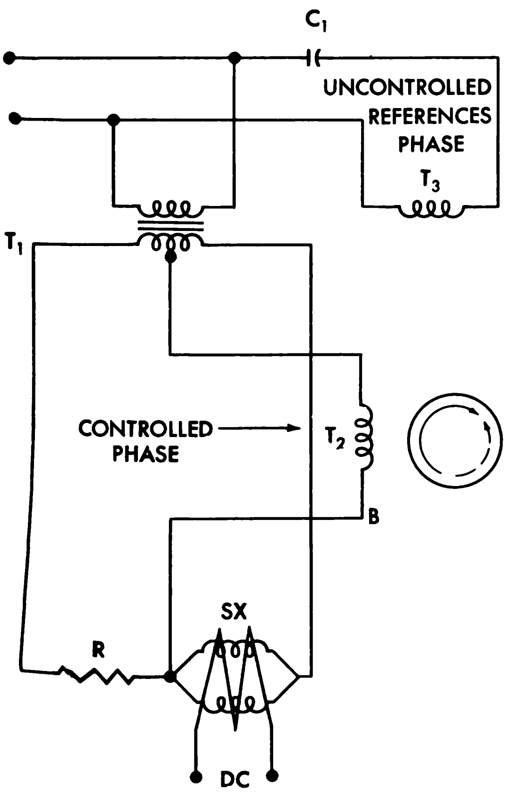
In conclusion, by varying the amount of DC control current, we can cause the load voltage DB to be nearly in-phase with the AC applied voltage or almost 180 degrees out-of-phase. It is also possible to get any phase relation between these extremes as indicated by the locus of the vector DB.

The output of a saturable reactor circuit can be varied in amplitude and phase; therefore, its output may be used to drive a servo

motor with variable torque in either direction. Hence, a high torque motor or indicating device can be driven and controlled with very small changes in DC control power. For this reason, the saturable reactor is often called a magnetic amplifier.

MAGNETIC AMPLIFIERS

Actually, a magnetic amplifier is an electrical power amplifier utilizing one or more saturable reactors together with combinations of rectifiers, resistors, and transformers which are necessary to utilize the amplifying properties of the transductor. The fact that no moving parts and no vacuum tube elements are necessary is one of the most important advantages of the magnetic amplifier. Another advantage of magnetic amplifiers is their freedom from deterioration, since all components can be designed to be practically everlasting.



Nonpolarized Magnetic Amplifier for Servo Control

Magnetic amplifiers may be contained in a sealed case; they are highly efficient, rugged, simple, and yet very reliable. For these reasons they are being used to replace vacuum tubes as power amplifiers in many control circuits.

Nonpolarized Magnetic Amplifier

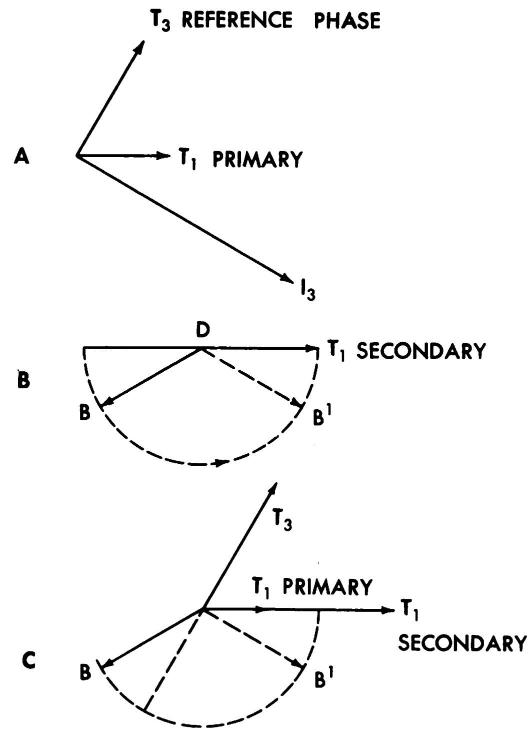
It will be recalled that it is necessary to invert only one of the AC fields to reverse the direction of rotation of a two-phase induction motor. The figure illustrates a method in which a saturable reactor phase-shifting bridge may control the direction of rotation and torque of a two-phase induction motor.

Note that the primary of the supply transformer T_1 is now connected across a capacitor and coil in series. The coil T_3 provides an uncontrolled phase of the induction motor field, and T_2 provides the variable controlled phase. The size of the capacitor C_1 is selected to be nearly resonant with the coil T_3, so that its current I_3 will lag the applied voltage of the primary only slightly as shown by the *Servo Motor Field Vectors* at A in the next illustration. Notice that this establishes a fixed reference for the uncontrolled phase T_3 which leads the primary applied voltage T_1 by nearly 90 degrees.

We have seen previously how a small change in DC control current—zero to a few milliamperes—can swing the T_2 voltage vector (DB) counterclockwise along its circular locus from point B to B' as indicated by the accompanying vector at B.

Now consider the no DC control current (unsaturated) condition, where the controlled phase of the motor field T_2 (vector DB) may be represented by position B. Here we can see that the controlled field (phase T_2) leads the reference field (phase T_3) by approximately 90 degrees. This satisfies the requirements for rotation of the induction motor—say in the clockwise direction.

Now assume there is enough DC control current flowing to saturate the reactor S_X. Under this condition, the controlled field will



Servo Motor Field Vectors

swing counterclockwise to position B', shifting its phase by nearly 180 degrees. The accompanying vector at C illustrates this 180-degree phase shift of the controlled field from position B to B'. This shows that the controlled field now lags the reference field by approximately 90 degrees. This satisfies the requirements for rotation of the servo motor in its reverse—counterclockwise—direction.

If the DC control current is such that the reactor is partially saturated mid-way between these two extreme conditions, the controlled field would be 180 degrees out-of-phase with the reference field. The servo motor would not be energized properly to produce torque, resulting in zero rotation.

Polarized Magnetic Amplifier

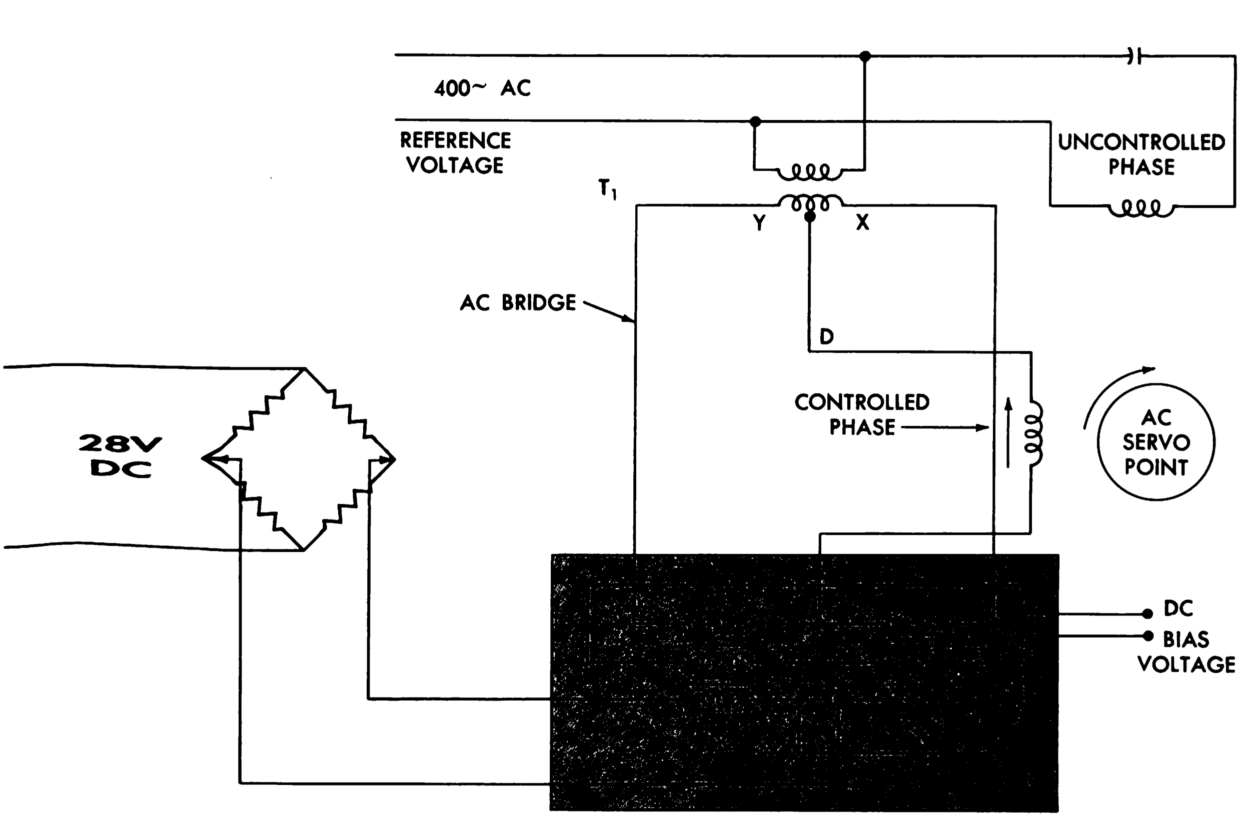
The simple arrangements discussed up to this point consisted of nonpolarized magnetic saturable reactors which were unable to discriminate between positive and negative DC currents in the control circuit. On the other hand, a magnetic amplifier utilizing a polarized saturable reactor is influenced by changes in direction of the DC control current by using separate control coils ($N_B{}^1$ and $N_B{}^2$) biased with a particular polarity reference as indicated in the next illustration.

In this *Polarized Magnetic Amplifier*, each saturable reactor (SX_1 and SX_2) has three windings, providing the two lower legs of an AC bridge. It will be noted that the DC bias current flows through a winding on each reactor and that these windings ($N_B{}^1$ and $N_B{}^2$) are connected series-aiding. On the other hand, while the DC error control current also flows through a winding on each reactor, these windings ($N_C{}^1$ and $N_C{}^2$) are connected series-opposing.

With the DC error bridge balanced, there will be no DC error control current, and the two reactors (SX_1 and SX_2) should be equally and partially saturated by the DC bias current. The two reactors being equally saturated results in equal reactance values for the two AC coils $N_L{}^1$ and $N_L{}^2$, providing a balanced condition for the AC bridge. Points D and B are at the same potential; therefore, there can be no AC current through the controlled phase of the servo motor field, and the servo motor is not energized.

Let us now assume that the DC bridge, controlling the error current, is unbalanced, creating a DC error signal with the polarity indicated. It can be seen that the error signal current flowing through the control winding $N_C{}^1$ tends to oppose or cancel the effect of the DC bias current through $N_B{}^1$, thereby decreasing the saturation of the reactor SX_2 and increasing the reactance of its AC winding. Conversely, the same error signal current flowing through $N_C{}^2$ acts to aid the DC bias current through $N_B{}^2$ and tends to further saturate the reactor SX_2.

This increased saturation produces a considerable decrease in the reactance for the



Polarized Magnetic Amplifier

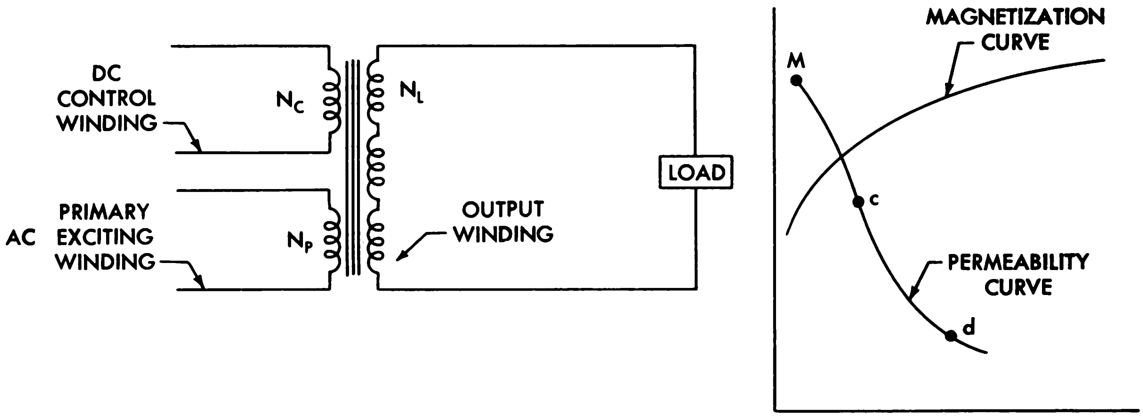
AC winding of SX_2, effectively connecting point B to point X which places the controlled phase of the servo motor field effectively across the right-hand half of the secondary supply voltage. This establishes one polarity for the controlled field, energizing the servo motor in such a way as to cause clockwise rotation as indicated by the solid arrow in the illustration. Reversal of the error signal polarity causes the system to operate in the same manner but in the opposite direction.

Saturable Transformer Magnetic Amplifier

The permeability of a transformer core may be altered by the manner in which the permeability of a saturable reactor is varied. In the case of a saturable transformer, changing the permeability or degree of saturation of the core material changes the number of lines of induction between the primary and secondary windings. Hence, the voltage induced into the secondary by an alternating current of fixed value in the primary will depend upon the permeability of the core or the effectiveness of the core in producing coupling between the primary and the secondary.

Consider the *Saturable Transformer* illustrated next where an AC voltage is applied to the AC exciting winding and a load is connected across the output winding. A third winding, called the control winding, is also shown.

Varying the DC current in the control winding varies the magnetizing force through the core. This, in turn, causes its permeability to vary in the manner shown in the accompanying graph. With highly permeable substances, the region *mcd* of the permeability curve is extremely steep. This means that a small change in magnetizing



Saturable Transformer and Its Characteristic Curve

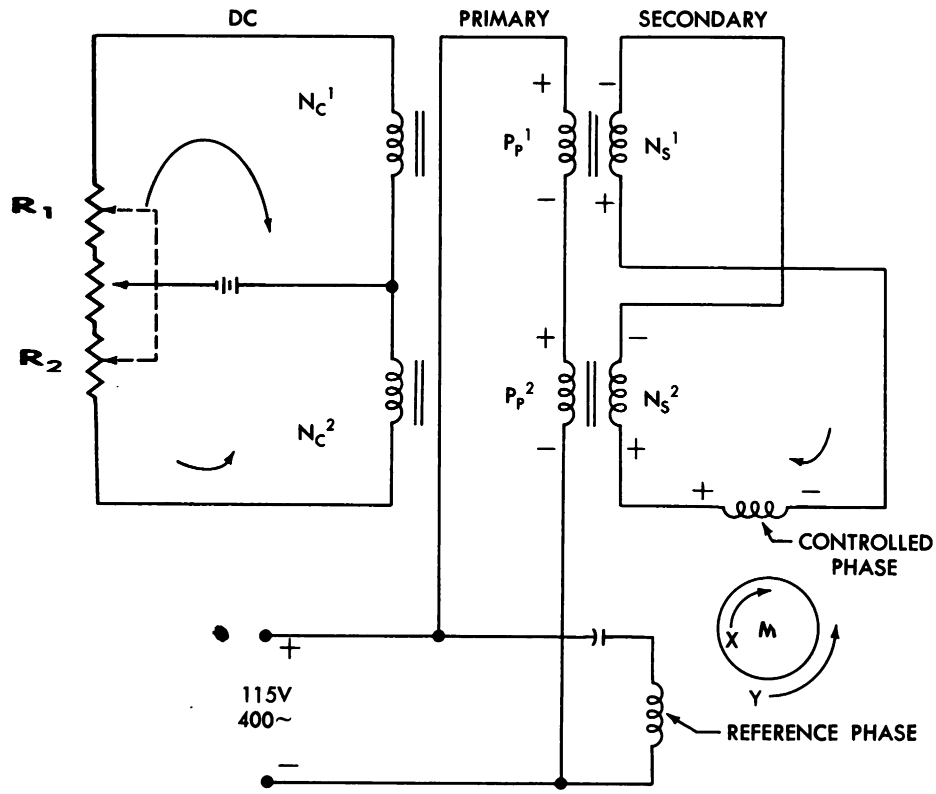
force causes a large change in permeability. In other words, a small increase in DC control current increases the degree of saturation, producing an almost constant flux density, and there can be no induced voltage with a stationary flux. However, a small decrease in DC control current decreases the degree of saturation, allowing a considerable change in flux density to produce a large relative motion between flux and conductor and hence a large induced voltage. Therefore, a small change in DC control current causes a large change in the output voltage from the secondary winding. Large amounts of power may thus be controlled by small amounts of power in the control winding. In other words, the saturable transformer provides high power amplification.

The basic circuit just described is only a portion of the *Saturable Transformer Magnetic Amplifier* illustrated next containing two output transformers that are completely insulated from each other. Notice, in this circuit, that the secondary windings $N_S{}^1$ and $N_S{}^2$ are wound so that their individual induced voltages will at all times be series-opposing.

For explanation purposes, let us assume that the control potentiometer, shown in the illustration, is balanced; that is, the movable contact is adjusted to the midposition between the two equal resistances R_1 and R_2. As a result, R_1 and R_2 conduct equally, producing an equal current flow through both DC control windings $N_C{}^1$ and $N_C{}^2$ and causing both sections of the saturable transformer to operate at the same point on the permeability curve within region *mcd* as previously illustrated. This, in turn, results in equal voltages being induced from the primary excitation windings into both secondary output windings. However, due to the fact that the voltages induced into the output windings are series-opposing as well as equal in magnitude, the net output voltage is zero. Under this condition there will be no current through the controlled phase of the servo motor field to produce torque; therefore, the motor will not rotate.

Now assume the control potentiometer is moved upward to position X, decreasing the R_1 component of resistance and increasing the R_2 component. This results in an increased DC control current flow through the upper control winding $N_C{}^1$ and a decrease in control current through the lower control winding $N_C{}^2$.

The increased current through the upper control winding of the saturable transformer increases its degree of saturation, decreases the effectiveness of its iron core to produce good coupling, and causes the voltage induced into the upper output winding to decrease. At the same time, the de-



Saturable Transformer Magnetic Amplifier

crease in current through the lower DC control winding decreases the degree of saturation of the lower saturable transformer, increases the effectiveness of its iron core to produce good coupling, and causes the voltage induced into the lower output winding to increase.

The resultant output voltage applied to the controlled phase of the induction motor field is the algebraic sum of these two induced voltages and will be of the same phase as the larger of the two induced voltages. In this case the output will be in-phase with the signal voltage in the lower output secondary winding as indicated. The amplitude of this output signal is determined by the difference between the two voltages induced into the two output windings which, in turn, depends upon the magnitude of the DC current in the DC control windings.

The phase of the controlled field will be inverted 180 degrees, and the direction of rotation will reverse if the control potentiometer is lowered toward position *Y*. The reason for this is that the voltage induced into the upper output winding would increase due to a decrease in current through the upper DC control winding, and the voltage induced into the lower output winding would decrease due to an increase in current through the lower DC control winding. As a result, the resultant output voltage would be in-phase with the signal induced in the upper output winding.

The preceding discussion of amplifier units of computers is not, of course, a complete discussion of all types of amplifiers used in interceptor aircraft systems. The intent here has been to cover only the less conventional types.

SECTION F, DIGITAL COMPUTERS

Since analog computers do not possess the ability to handle complex data rapidly, it has been necessary to develop special computers for specific applications. It has also been required to present data to the computer in some form which will permit rapid evaluation with a high degree of accuracy. These requirements have led to the development of a high-speed digital computer operating on binary principles. Such machines require only the ability to count to two and will accept information with a minimum of processing.

A very important difficulty in conjunction with digital computers arises from the inequality between the abilities of the machine and the operator who must run it. The machine must handle large amounts of data and work in minute detail. The *arithmetic unit* of a digital computer may be exceedingly fast, but it is so simple-minded that it must be instructed in great detail. However, the human being who is presenting the problem to the machine is much slower and has limited capacity for storing data. Fortunately, contemporary digital computers are sequence controlled or automatic, leaving only the initial planning or *programming* to the operator. The problem is that the instructions must be made absolutely complete and unambiguous in every detail, and yet be expressed in terms that the computer can understand.

A great deal of effort is saved in preparing instructions for digital computers when frequently used sequences, known as *subroutines*, are incorporated into the equipment. Ideally, a change in instructional requirements should be capable of being handled merely by the replacement of one punched card which can be quickly prepared or even prepared in advance. This is somewhat like what is done when a complicated plan for interceptor action has to be put into effect at a moment's notice. That is, the order may be reduced to "Operate plan two" or any simple coded order, so long as the details have been worked out in advance with perhaps the use of a flight simulator as discussed in the introduction to this chapter.

From the military point of view, there is the problem of working with extremely noisy coded telemetered signals and echo pulses from both acquisition (target tracking) radar and interceptor tracking radar. Hence, although the human being does have the ability to draw unprogrammed conclusions from data presented to him by computing equipment, he can not sample data and decide by looking at it whether it is good data or whether it is noise. And, of course, the enemy is constantly trying to make the data noisier by means of countermeasure jamming techniques. Sampling data must be done statistically, and it appears that the best manner of accomplishing this is on a digital statistical basis.

At the present time of the computing art, analog and digital techniques supplement each other well. Analog equipment is usually simple, and for many situations its precision is greater than the accuracy of its programmed input data. Digital type computers perhaps have greater flexibility and are used more extensively for *reiterative* solutions and nonlinear mathematics. They are used where self-checking features are desired, enabling them to recognize when they have made an error, allowing them to retrack and recompute the problem, and retaining in storage all partial results that had accumulated before the error was made.

Digital equipment is also utilized when the problem setup must be changed radically in the course of the problem or where sorting and other manipulations of data are necessary.

The present trend seems to be a combination consisting of a simple analog com-

puter to permit a problem to be understood and a digital computer for detailed investigations in the areas of a problem that require it.

DIGITAL COMPUTER TYPES AND COMPOSITION

You will find, as you progress through this section, that digital equipment is essentially composed of electronic, magnetic, or mechanical logic-switching circuits and counters which register and perform arithmetic and/or logic operations in separate steps.

Desk type calculating machines are perhaps familiar forms of the electromechanical type digital computer. The first automatic sequence-controlled computer used mechanical counters controlled by magnetic clutches and sequence relays. However, with the use of vacuum tubes for switching, storage, and arithmetic operations, a considerable saving of time for the accomplishment of these functions has been effected over similar equipment utilizing electromechanical mechanisms. Electronic digital computers had their beginning in the Electronic Numerical Integrator and Calculator (ENIAC). This computer was one of the first to use electronic circuits as the actual computing elements. It was designed primarily for step-by-step numerical integration of ballistic equations and for generating trajectories of projectiles and bombs. The ENIAC was designed and developed by the University of Pennsylvania, and it contains some 18,000 vacuum tubes.

It seems that the most common source of trouble in the ENIAC was the failure or short life of its much used element—the vacuum tube. The units of the ENIAC underwent a thorough routine operational check each week, and at one time—shortly after the machine was placed into operation—it was reported that about 250 vacuum tubes (actually bad) were replaced each week for several months.

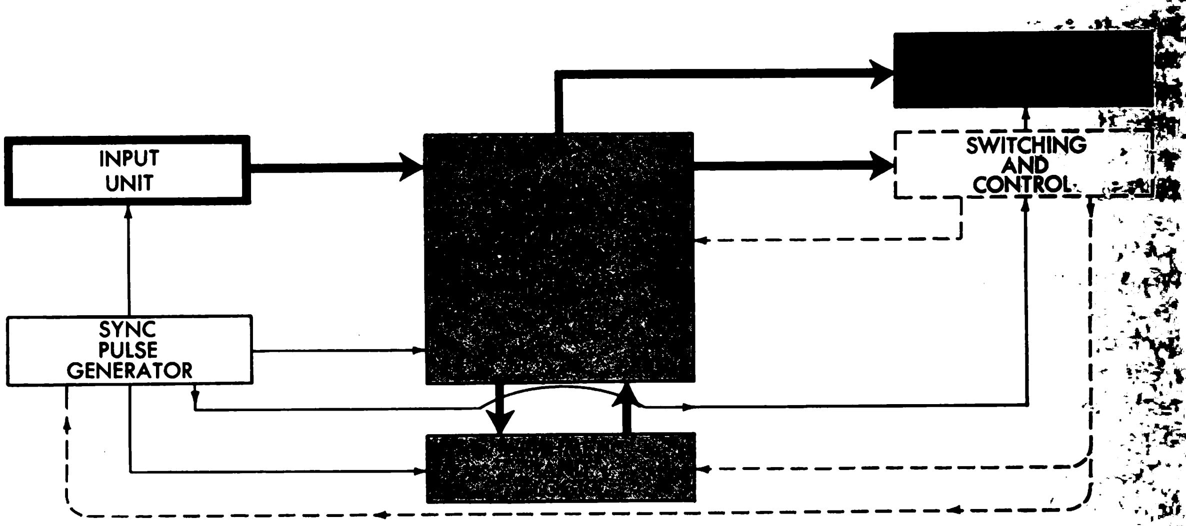
Today more and more computer units are made up of transistors, magnetic cores, and other solid state electronic devices for switching, storage, and arithmetic operations. Hence, a saving of time, space, weight, and power has been effected along with increased reliability. This has allowed rapid progress in the airborne and ground-based computer field. One of the first digital computers built making use of the newer techniques made available by these solid state devices is claimed to give a 50% reduction in size and a 90% reduction in power consumption.

Basic Functional Block Diagram

Only the general organization and function of the individual black-box units contained in the makeup of a basic digital computer will be presented here. Each block, as shown in the functional block diagram of a *Basic Digital Computer*, will be treated as a subject of its own and investigated more comprehensively in subsequent portions of this section.

All blocks shown consist of equipment composed of similar elements, but are interconnected to perform different functions. You must bear in mind, however, when studying the following paragraphs that, while the units do perform the various operations indicated, these operations may not be physically distinct. That is, the same components or types of components may be used for more than one function. For example, a component that performs arithmetical operations may also be called upon to perform a memory or storage function.

MAIN MEMORY STORE. Computers operating on the binary numbering system perform mathematical operations on numbers—expressed in the form of digits—to yield answers which are also expressed in digits. When input information is to be processed in digital form, a computer must be able to accept the information bit-by-bit and store each bit until all information for processing has been received. This requires that the digital computer have a memory or storage



Basic Digital Computer

unit as is illustrated in the functional block diagram of a *Basic Digital Computer.*

The information and instructions in the form of digits are retained in this storage unit until the computer is called upon to use the information to complete the problem. The type of memory or storage device used will determine the length of time that information can be retained.

DATA TRANSMISSION. The memory unit, like the human brain, requires some means of transferring information to and from various units of the computer system. These lines, links, or channels by which information is transferred from one unit to the other are called *data-transmission systems.*

In actual equipment you will find these links to be both electrical and mechanical. It has been said that a computer data-transmission system corresponds to the nervous system in the human body.

LOGIC UNIT. The function of the logic unit illustrated in the basic block diagram is to perform the necessary arithmetical processes in the proper sequence. This unit must be able to add, subtract, multiply, divide, extract square roots, and determine the proper sign (positive or negative). It must be able to transfer the results of the arithmetical processes through proper channels for storage or direct output, depending on instructions from the switching and control unit.

SWITCHING AND CONTROL UNIT. A switching and control unit must be included in a digital computer to instruct the equipment when to perform an operation and to govern the operation or sequence of operations to be performed. The control unit contains logic-switching circuits which are usually referred to as *AND, OR,* and *IF* connectives. These connective or switching circuits are activated by a program of instruction signals transmitted to them from a section of the main memory unit and from synchronizing clock pulses from the sync-pulse generator.

The sync-pulse generator, a kind of electronic distributor system, is used to produce sync-pulses which have a predetermined repetition frequency with a particular sync pulse period. These sync-pulses are applied simultaneously to a series of gates whose outputs synchronize all units to produce proper pulses at the proper point and at the proper time.

The units which receive the data and supply the answers are shown as the input and output units respectively. These units contain either photoelectric or electromechanical

equipment for reading-in and writing-out the desired numerical information—usually on punched cards, punched tape, or magnetic tape. Input and output units quite often employ temporary delay or storage systems composed of trigger circuits to hold portions of the programmed or output information while reading and writing is being accomplished.

Finally the computer requires motors, mechanical equipment, and electrical power or mechanical energy for its operation.

Before investigating the component parts utilized in the various subunits contained in the functional block diagram, it may be well to first discuss a few of the basic principles of logic on which digital computer operation depends.

BOOLIAN ALGEBRA

In order to operate properly and use some numbering system such as the binary system, a computer must have the ability to:

1. Store and remember data presented to it.
2. Make a choice based on previous results.
3. Perform long chains of operations.
4. Determine if the answer is correct.
5. Determine when one problem is finished and when to start on the next.

The ability of a computer to perform these functions will depend upon its design and the information supplied to it. As previously mentioned, an important requirement is that it be able to determine the answer to any problem by what is called logical truth. Logical truth differs from ordinary truth in that it involves both facts and suppositions based on facts. For example, that sugar dissolves in a teaspoon of water is an ordinary truth, but certain conditions are understood. That is, the amount of sugar is much smaller than the amount of water. If you attempt to mix a whole bag of sugar with a teaspoonful of water, you find that it would be impossible to dissolve all of the sugar. A computer cannot comprehend such limiting conditions. Certain conditions must be supplied so that it may form a logical pattern of operation. An example of a logical pattern of reasoning is expressed as follows:

1. All people are animals.
2. All animals are mortals.
3. Therefore, all people are mortals.

We often call such statements logical *syllogisms;* however, we would like to call your attention to the fact that all syllogisms are not necessarily logical. For example, in the following statements you can see that statement *3* is not true even though *1* and *2* are true:

1. All animals have legs.
2. All people have legs.
3. Therefore, all animals are people.

In everyday life, you use many logical truth patterns without realizing that you are doing so. Most of our simpler logical patterns are distinguished by words such as *AND, OR, IF, ELSE, NOT,* and *THEN.* In the language of mathematics these logical patterns may be indicated by *PLUS, MINUS, TIMES, DIVIDED BY,* or combinations of these.

When a set of statements is made, we can set up a table that will determine the truth values of our statements. In setting up this table, we will find it necessary to use some of the logical connectives such as *AND, OR,* and *IF.* Finally, we could have a table that would list the truth or falsity of each of the conditions of the statements we were considering.

Consider the simplest means for giving information—an answer in terms of either *yes* or *no.* Most facts can be defined by answers of *yes* or *no* to a series of questions. Hence, we can measure information

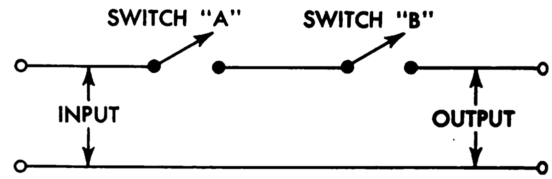
necessary to convey a fact by the resulting sequence and the number of terms *yes* and *no* that are required for the fact in question.

This system of two alternatives is known as the binary system, and the information is represented by the words *set* and *unset*—meaning conducting and nonconducting. Subsequently, we will investigate how this binary principle is applied to the numbers *1* and *0*, and how these numbers are represented by the presence or absence of a pulse at a particular instant.

The process of mathematical logic—setting up mathematical relations of truth values of statements and performing mathematical operations on these statements—is called Boolian Algebra. This process was so named because it was first introduced by an English mathematician, George Boole, in his book *The Laws of Thought,* published in 1854. You can actually apply this algebra of logic to the study of electrical circuits and switching, and by means of electrical circuits and switching, you can study the algebra of logic.

By using certain types of circuitry, we can determine mathematically the truth values of such statements as those that would imply connectives such as *AND, OR, IF* and *ONLY*. Each of these connectives may be represented by the conducting conditions of various switching circuits and designated by a number. For example, if we let the truth of a statement be shown as the digit *1*, and the falsity of the statement as the digit *0*, we can determine certain facts concerning a series-switching circuit. Take a look at the *Series Switching Circuit* illustrated and notice that, if either switch is open, it will be indicated in the accompanying table as *0*—no current flowing. On the other hand, if either switch is closed, its position is represented by the digit *1*. However, you will note that only one condition of the circuitry allows current to flow in the output—that is, when both switches are in closed position.

This condition simulates the truth values of statements which use the connective *AND*



SWITCH "A"	SWITCH "B"	OUTPUT
OPEN 0	OPEN 0	NO CURRENT 0
CLOSED 1	OPEN 0	NO CURRENT 0
OPEN 0	CLOSED 1	NO CURRENT 0
CLOSED 1	CLOSED 1	CURRENT FLOW 1

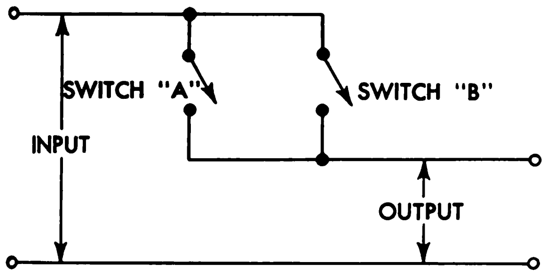
Series Switching Circuit

because, for current to flow, switch *A* and switch *B* must both be in the closed position representing *truth* and symbolized by the digit *1*. No current will flow if either or both switches are open; hence, the open position represents *false* and is symbolized by the digit *0*.

To determine the truth value of statements using the connective *OR*, circuitry similar to the *Parallel Switching Circuit* shown next may be utilized. In this circuit, you can see that current will flow in the output circuit when either switch is closed or when both switches are closed. No current will flow when both switches are open. Thus a table of truths for the parallel switching circuit may similarly be set up as shown in the accompanying illustration. In this manner a computer can scan the truth values of statements containing a variety of logical connectives and may be used to indicate in digital form the truth or falsity of a statement.

Binary Representation of Numbers

You will find that the binary system of counting and coding appears frequently in automatic digital computer equipment because only *two* symbols are required in



SWITCH "A"	SWITCH "B"	OUTPUT
OPEN 0	OPEN 0	NO CURRENT FLOW 0
CLOSED 1	OPEN 0	CURRENT FLOWS 1
OPEN 0	CLOSED 1	CURRENT FLOWS 1
CLOSED 1	CLOSED 1	CURRENT FLOWS 1

Parallel Switching Circuit

binary notation—*1* and *0*—as opposed to the ten characters—*0, 1, 2, 3, 4, 5, 6, 7, 8,* and *9* —in the decimal system.

Because only two symbols are required in the binary system, they can be represented electrically by a voltage being either present or absent at some point in a circuit. They may be represented by a switch being either opened or closed or by any bistable (two-state) device such as a multivibrator trigger circuit or gate.

Relays, vacuum tubes, magnetic cores, and transitors can readily distinguish between *on* and *off* conditions. All of these elements may be pulse-biased beyond cutoff to represent an *unset, off,* or *nonconducting* condition. Pulses of the proper amplitude and polarity may also be applied to those elements so that they may represent a *set, on,* or *conducting* condition.

The punched tape or card systems used to program instruction data into computers and to record their outputs are also very suitable for use in equipment employing binary principles. For example, the presence of a hole in a card may represent the *on* or *set* condition, and the absence of a hole could represent the *off* or *unset* condition.

No doubt you are familiar with the usage of the scale-of-ten (decimal) representation of numbers. For example, the number *6,501* may be expressed in powers of ten in the following manner in which all digits (*0* through *9*) may be required to express a number:

$$(6\times10^3)+(5\times10^2)+(0\times10^1)+(1\times10^0)$$

$$6000 + 500 + 0 + 1$$

In the binary representation of a number, we need only two digits, *0* and *1,* in various combinations. A sequence of *0's* and *1's* in the binary system indicates the presence of the number *2* raised to some power. For example, the digits *1101* in the binary system, representing the decimal number *13,* may be translated into decimal notation as shown below.

Binary to Decimal Conversion

$$1 \quad 1 \quad 0 \quad 1$$

$$(1\times2^3)+(1\times2^2)+(0\times2^1)+(1\times2^0)$$

$$8 + 4 + 0 + 1 = 13$$

On the other hand a decimal number may be converted to binary notation using the reverse of the above process—that is, by a repeated division by *2* as shown next.

Decimal to Binary Conversion

$$\frac{13}{2} = 6 \text{ with remainder of } 1$$

$$\frac{6}{2} = 3 \text{ with remainder of } 0$$

$$\frac{3}{2} = 1 \text{ with remainder of } 1$$

$$\frac{1}{2} = 0 \text{ with remainder of } 1$$

First the decimal number *13* to be translated was entered into the first column. The division *13/2* was performed, and the quotient *6* was placed directly below with the remainder entered in the last column. This process is repeated, as indicated, until the resulting quotient is no longer divisible by *2.*

The binary representation of the decimal number *13* is obtained by copying the remainder column, starting with the *last remainder* (representing the first digit) and proceeding upward. Hence, the binary equivalent of decimal *13* is *1101*.

Binary Numbering System and Decimal Equivalents

Decimal	Binary	Decimal	Binary
0	0	17	10001
1	1	18	10010
2	10	19	10011
3	11	20	10100
4	100	21	10101
5	101	22	10110
6	110	23	10111
7	111	24	11000
8	1000	25	11001
9	1001	26	11010
10	1010	27	11011
11	1011	28	11100
12	1100	29	11101
13	1101	30	11110
14	1110	31	11111
15	1111	32	100000
16	10000	33	100001

In the accompanying table you will find decimal numbers *0* to *33* expressed in their equivalent binary notation. Examine this table for a moment and consider the first four decimal numbers *0, 1, 2,* and *3*. If these numbers are to represent four different conditions, you will find that the use of the decimal numbering system in a computing device would require four separate registers to record the information presented by the four numbers. However, comparing the decimal column to the binary column, you will see that the binary system needs only two digits, *0* and *1*, to represent the first four numbers. Thus, we can represent *four* conditions of information by the use of *two* digits.

Now let's see how an electronic on-off system can be adapted to the presentation of data in binary form. Suppose that we would like to indicate the decimal number *3* in a computing operation. In a binary digital computer it is represented as *11*. These binary digits could be represented in an electronic system by two relays both of which are in a closed (set) condition. Also the number *2*, represented in the binary system by *10*, could be indicated by the first relay in a closed (set) condition and the second relay in an open (unset) condition.

Once more referring to the table, you will notice that three digits of the binary system may be arranged to represent the numbers *0* through *7* in the decimal system or eight *bits* of information. Similarly, with four binary digits we can represent a total of *16* bits of information; and with five binary digits, *32* bits of information. For example, the number *31* (binary *11111*) would require five relays all in the closed (set) condition, the number *32* (binary *100000*) would require six relays with the first relay closed and the remaining five open.

From the previous paragraphs we have seen that the binary system has distinct advantages for electronic data processing equipment. Reducing the number of figures or distinguishable states from *ten* to *two* seems to be as tremendous simplification, even though it necessitates more digits per numerical representation. To stress the case in point, large decimal numbers are handled easier with *two-way* switches than with *ten-way* switches.

Thus far, we have presented the requirements and principles involved in an electronic device that will make accurate computations in very short periods of time. [illegible] now investigate some of the actual devices and some of the basic circuitry within such a computing system. The circuits will be considered from the overall standpoint of their possible uses, as it would be impractical to try to discuss all of the applications of any one device or to state positively that one particular device would satisfy the requirements of a problem better than some other device.

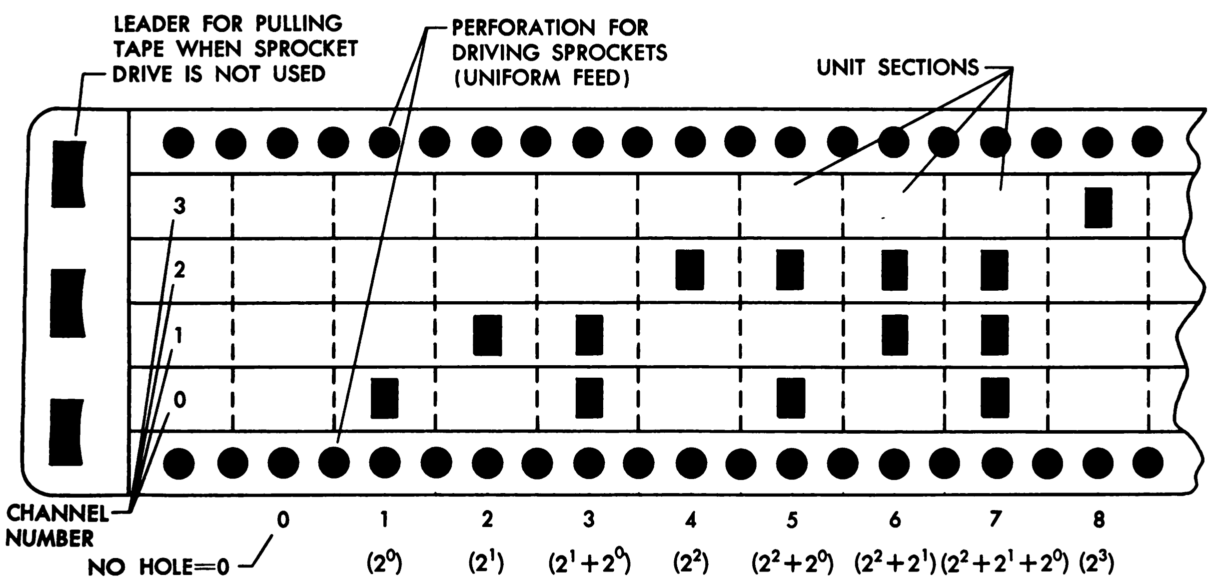
Input Units

The input unit is that part of the computing system into which all the known data (statements) of a problems are inserted. This data is usually supplied to the system by means of punched paper tape or cards, photographic tape (film), magnetic tape, or wire upon which the data and instructions have been recorded.

PUNCHED TAPE. The figure illustrates a typical *Four-Channel Punched Tape with Binary Coding.*

Notice the similarity to an ordinary strip of movie film. The original information is transferred to the tape in accordance with a pattern that will represent certain digits to be used in the computer. In the case of the binary digital computer, these digits will be *1* and *0*, with the punched holes representing *1* and the absence of holes representing *0*. In a typical binary punched tape system, the tape is divided into a number of channels. The position of a punched hole, as determined by the channel in which it occurs, indicates the number *2* raised to a power corresponding to the number of the channel.

For example, you will notice that a hole punched in the zero channel (at the bottom of the tape horizontally) represents 2^0 or *1*, and so on. In all cases, the absence of a hole in any unit section of a channel represents *0*. Thus, the decimal quantity *7* is indicated by holes in channels zero, one, and two representing the sum of $2^0+2^1+2^2$, or $1+2+4$. Now look at the accompanying table *Comparing Channel Numbers with Binary Digits,* locate the number *7*, and notice how this particular position on the punched tape indicates the binary notation for the number *7*. A tape with four channels can represent *16* separate bits of information. With additional channels, a considerably larger amount of information can be handled. As you can see, decimal quantities may be represented by a suitable arrangement of holes in the punched tape to provide input data for computers using the decimal or denary principle of operation.



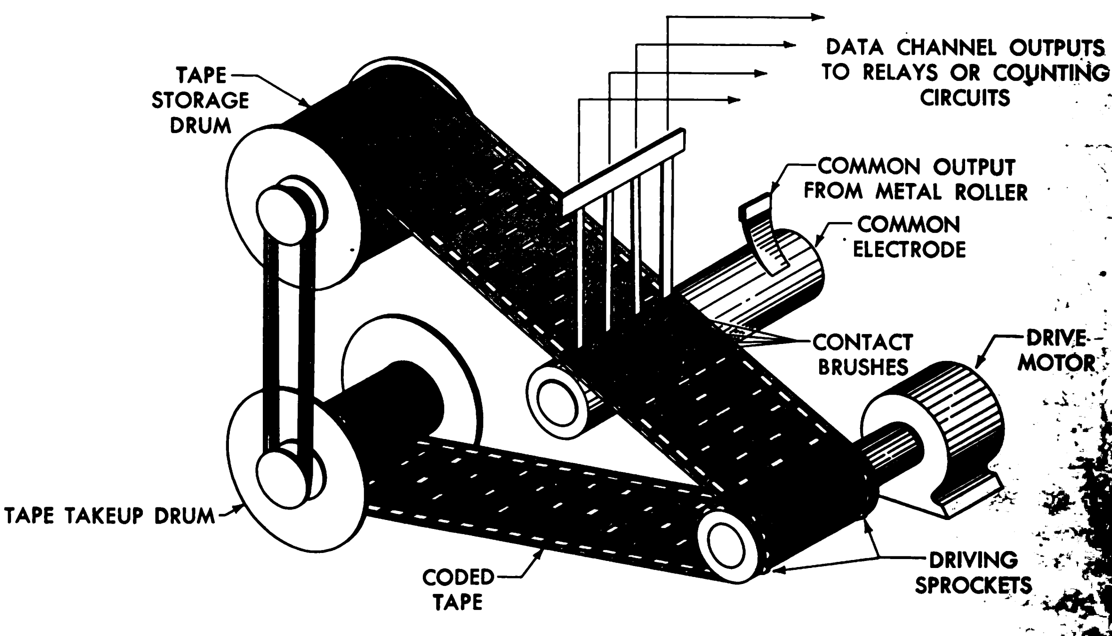
Four-Channel Punched Tape with Binary Coding

Comparing Channel Numbers with Binary Digits

Decimal Quantities	Channel Number	Binary Digits
	3 2 1 0	
0	0 0 0 0	0000
1	0 0 0 1	0001
2	0 0 1 0	0010
3	0 0 1 1	0011
4	0 1 0 0	0100
5	0 1 0 1	0101
6	0 1 1 0	0110
7	0 1 1 1	0111
8	1 0 0 0	1000
9	1 0 0 1	1001
10	1 0 1 0	1010
11	1 0 1 1	1011
12	1 1 0 0	1100
13	1 1 0 1	1101
14	1 1 1 0	1110
15	1 1 1 1	1111

A Punched Tape Reading Unit shown next illustrates how a punched tape is put to use.

Note the tape passes through the reading unit which permits electrical contact to be made between a metal drum or roller electrode and a brush whenever a hole passes between them. To prevent slippage and assure an even tape speed, the tape has a row of uniformly spaced perforations along both edges or one edge. The teeth of a sprocket driving gear turning at a constant speed are meshed into the perforations of the tape. However, the difficulty in maintaining uniform tape speed results from an increase in diameter of the take-up drum as the tape is wound upon it. This tends to cause a gradual increase in the rate at which the tape is unwound from the storage drum and necessitates compensation by a corresponding reduction of the driving motor speed or by other means. We can use the punched tape—or a similarly punched card—to sort



A Punched Tape Reading Unit

list, select, and copy information to make selections and comparisons according to specific instructions. In other words, the tape can be made to control mathematical operations such as addition, subtraction, multiplication, division, and extraction of roots.

You are perhaps familiar with the fact that, in the case of a movie film, sound patterns are recorded on the edge of the film. The patterns are then converted into sound by means of photoelectric cells. In a similar manner, we must convert the predetermined program data upon punched tapes or cards. This may be accomplished by means of a typewriter-like device which perforates the tape or card rather than printing symbols upon its surface. This information can then be converted into electronic pulses by means of the contacts shown in the illustration or by means of photoelectric cells.

MAGNETIC TAPE OR WIRE. Magnetic-tape input equipment resembles a magnetic sound recording machine. Usually the tape is stored on two reels and when being wound from one to the other, it passes across a *reading* or *writing* head as illustrated.

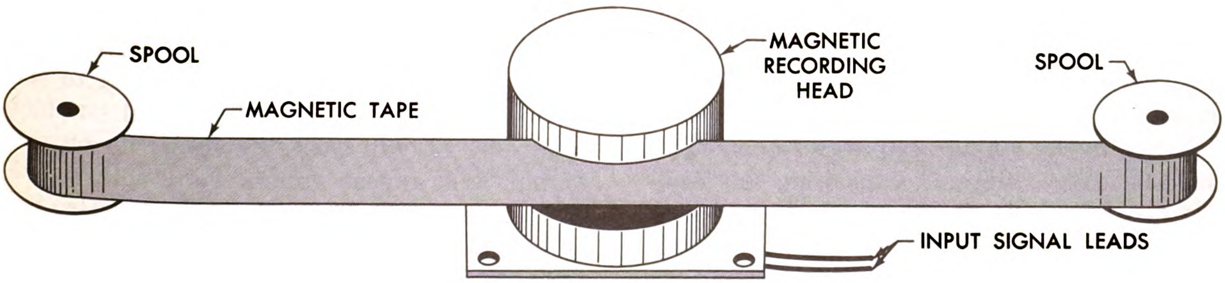
The pattern in which the digits are recorded is similar to that of the holes punched in paper tape, and several parallel tracks are used. However, in the case of magnetic tapes or wires, the input information is recorded in the form of magnetized spots or areas.

Variations in the degree of magnetization (which represent variations in recorded data) are used to produce changes in current flow in the circuits of the reading unit. These changes in current can be amplified and used to operate relays or other devices within the computer.

In addition to the recorded types of information that may be supplied to the input unit of a computer, it is possible for the operator to initiate the instructions to the computer and apply them through the use of push-button switches or keys.

DECODING. In airborne interceptor equipment, the computer may receive data in the form of analog error signals from sensor and reference units. These signals may be mixed in a certain ratio or fed directly to a command computer in accordance with a predetermined program of coded instructions so that rapid guidance or attitude corrections can be accomplished.

Pilotless interceptor computers may receive initial instructions in the form of coded transmitted pulses from a ground control station. Information may be transmitted by separate or continuous changes in pulse repetition rate or by pulse-frequency modulation, pulse-time modulation, or pulse-width modulation. Hence, a very important function of the input units of airborne guidance computers is the coding and decoding of information pertaining to flight. Enemy countermeasures or guidance of more than one interceptor at a time may make this operation very necessary.



Magnetic Tape Recording

Analog to digital as well as *digital to analog* converters may be necessary to provide the input units of the computer with proper coded information. The process of coding and decoding is a very important function of the input units of computers. The information received must be converted into a form that will be suitable for the operation of other circuits or units.

Preparation of input information may be accomplished by demodulation of any of the types of modulation mentioned above. Discriminating circuits may be used to select pulses of a particular width, amplitude, frequency, phase, or time difference, and to reject all others. Interference in the form of noise is reduced to a minimum by these decoding circuits in order to develop reliable error signals.

NOTE

A more comprehensive treatment of the various methods of modulating and demodulating transmitted information, to and from computer units contained in airborne interceptors, will be presented in the following chapter—Interceptor Guidance and Space Stabilization.

A *Pulse Decoding Relay,* of the type illustrated, is especially suitable for use in digital computers. In this circuit, an input signal composed of narrow and wide pulses may be decoded.

For example, when narrow pulses with long spacing are received, relay A will close. The second relay B will not close since the pulse is too short in duration to permit capacitor C_b to charge through series resistor R to the voltage required for energizing the relay. Signals consisting of wide pulses with short spacing will charge both capacitors C_a and C_b which will cause both relays to close—with relay A tending to close slightly ahead of relay B. The variation in closing time may be compensated for by increasing the tension of the contacts of relay B.

This type decoder may also be utilized in the switching and control unit of data processing digital computers.

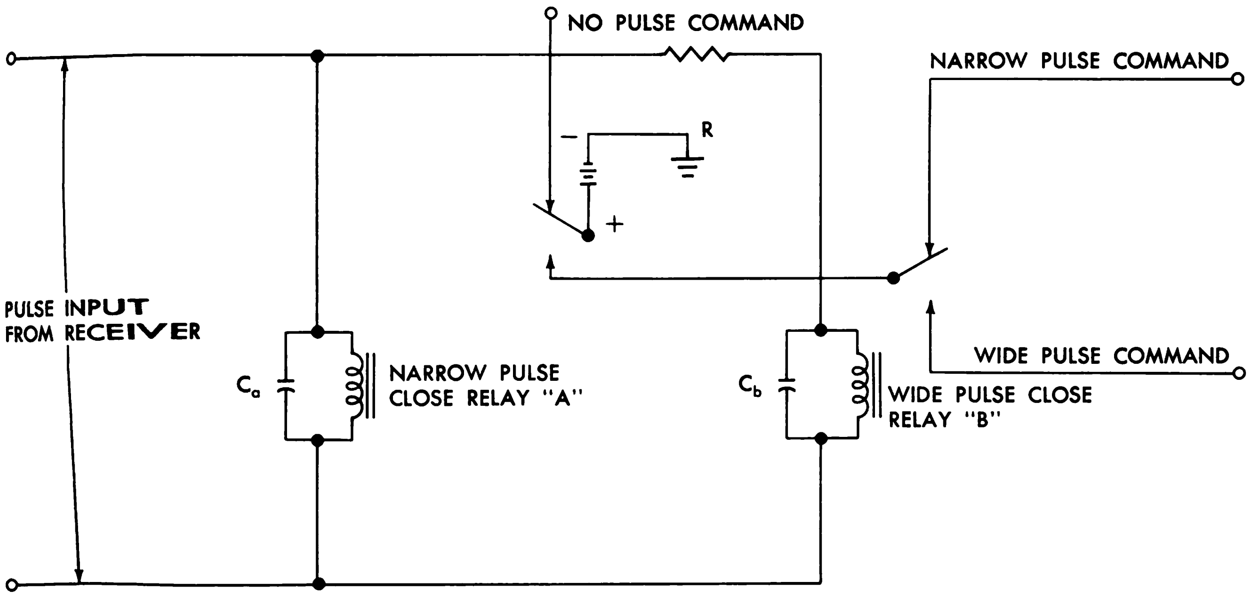
Arithmetic and Logic Units

The basic operations of any digital computer are dependent on ordinary arithmetic operations such as addition, subtraction, multiplication, and division. These arithmetic operations are mainly accomplished by means of two basic circuit types, both of which have two stable states—*on* and *off.* These two basic circuits consist of variations of the bistable multivibrator (flip-flop) circuit and variations of gating circuits.

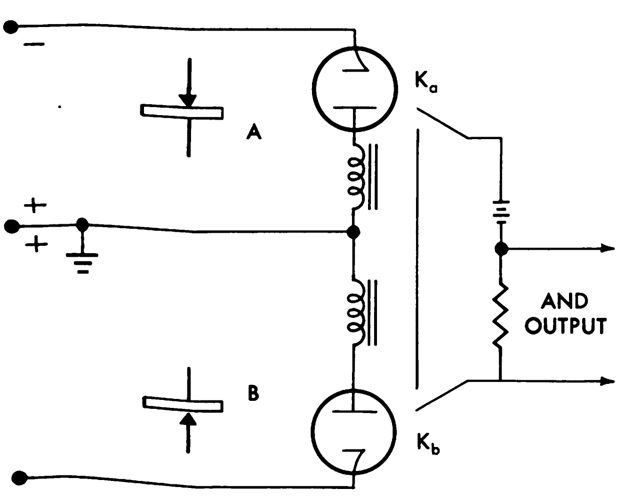
GATING CIRCUITS. Earlier in this section a switch was used as an example of a basic device that will readily represent the conditions of either *go* or *no-go.* Many devices have been developed to simulate the action of a switch and provide two stable states. Electronic components that provide these characteristics are referred to as *gates* or *gating circuits.* The gate, as its name implies, will permit or prevent the flow of current in a circuit.

An electric gate may consist of a circuit with a single output and two or more inputs. The circuitry of some gates may be so arranged that an output signal or pulse is produced when, and only when, input pulses are fed to a particular set of input leads simultaneously. Such circuits are also known as *coincidence circuits* or *LOGICAL-AND* circuits.

A *Fundamental Diode LOGICAL-AND Gate* is illustrated. When information in the form of a negative voltage is applied to the cathode of diode A, this diode will conduct and relay K_a will close. However, no output signal will appear unless both relays are closed. This means that diode B must also receive a negative voltage at its cathode simultaneously with that received by diode A. Thus, it may conduct and cause relay K_b to close. When diodes A and B are both con-



Pulse Decoding Relay



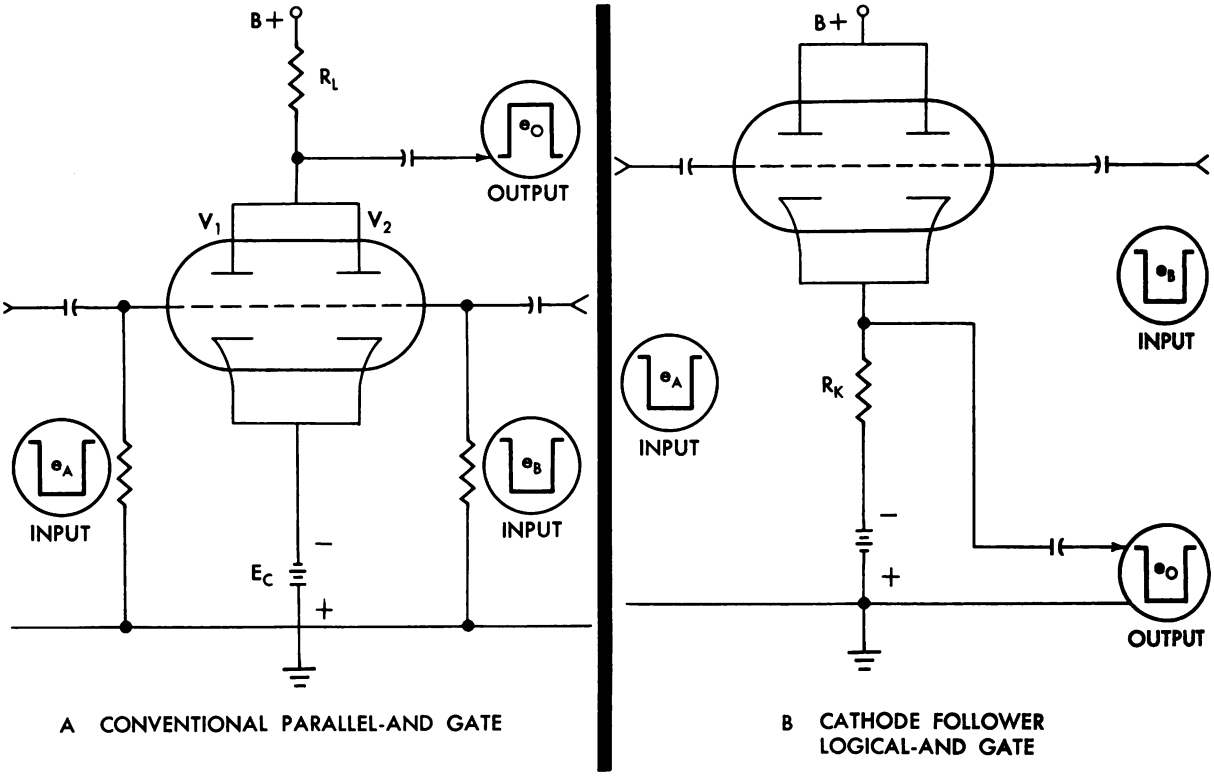
Fundamental Diode LOGICAL-AND Gate

ducting, current will flow in the output circuit and the connective *AND* is electronically simulated.

In addition to diodes, almost any multi-element tube may be used as a *LOGICAL-AND* gating device. For example, see the dual triode *LOGICAL-AND gate circuits.*

The *Conventional Triode LOGICAL-AND Circuit,* at A, consists of two triode sections contained in one envelope, the plates of which are tied together and connected to the plate supply through a common plate load resistor. Since both triode sections are biased positive with respect to their common cathodes, they are normally set in their conducting state. If only one triode section at a time is driven below cutoff by a negative input pulse, the total current flow through R_L will be virtually unaffected, because its parallel section remains in the state of saturation. Hence, the output voltage will rise only slightly if V_1 receives a negative input signal.

On the other hand, if negative trigger pulses are applied simultaneously to both sections—driving V_1 and V_2 below cutoff—neither section will conduct. Therefore, the voltage drop across R_L will suddenly decrease to zero, and the output voltage will rise abruptly to its maximum positive value. The output voltage will remain in this positive state so long as both inputs (e_A and e_B) maintain their respective sections (V_1 and V_2) below cutoff.



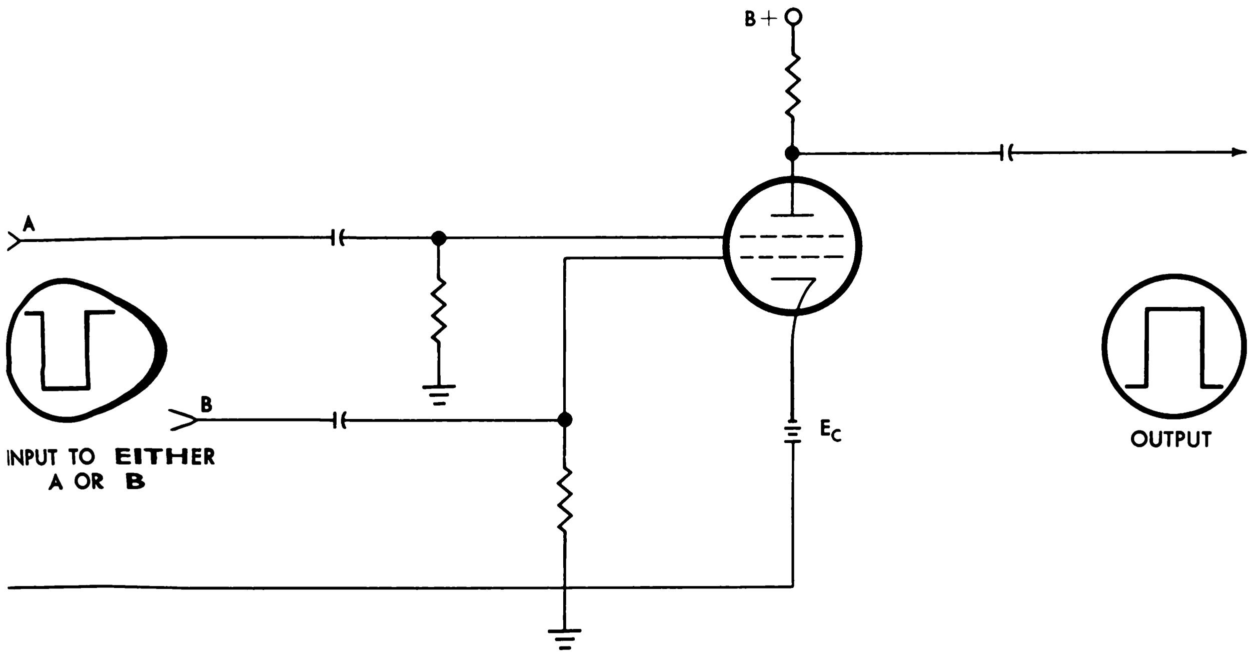
Dual Triode Parallel-AND-Gate Circuits

Note that the waveforms accompanying the conventional triode *LOGICAL-AND* circuit show the output (upon coincidence of two negative inputs) to be a positive pulse.

A *Cathode Follower LOGICAL-AND Gate* shown at **B** is similar in operation in that both sections are again biased in their conducting state. That is, during absence of both input pulses, the cathodes of both sections are maintained slightly positive with respect to ground. Here again, if only one triode section at a time is driven into cutoff, the change in total current flow through the load (now in the cathode circuit) is negligible. However, if negative pulses of the proper magnitude are fed to both grids in coincidence and both sections are cutoff simultaneously, cathode current ceases to flow and the output voltage abruptly decreases. Hence, in this circuit the output follows the input. That is, the output (upon coincidence of two negative inputs) is also negative as shown by the waveforms accompanying the illustration.

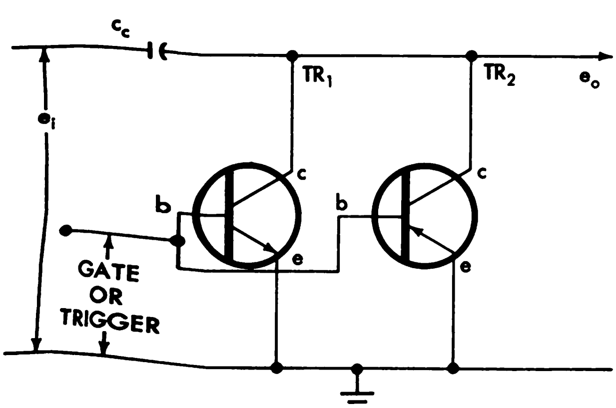
As you might expect, the cathode-follower gate has the desirable inherent characteristics of lower output impedance and good transient stability due to degenerative feedback. Also, because of its in-phase output, it lends itself well to the cascading of several such stages.

A *Tetrode LOGICAL-OR Gate,* illustrated next, supplies a positive output pulse if a negative trigger is applied to either grid—*A* or *B*. Since both grids in this circuit are normally biased positively, the tube is set in its conducting state. The tube will remain in this stable state until driven below cutoff by a negative pulse which may be applied to either grid *A* or grid *B*. Hence, if either input is pulsed, plate current will cease flowing, and the output or plate voltage will abruptly rise to its maximum value as indicated.



Tetrode LOGICAL-OR Gate

Transistor Gating or Switching Circuits seems to be making a fair bid to take the place of their electromechanical and vacuum tube equivalents. A *Basic Transistor Gate* making use of the *all or nothing* type action is illustrated in the accompanying diagram.



Basic Transistor Gate

Any signal (positive or negative) appearing at the common collector input by way of capacitor C_C will be shunted to ground by the conduction of either TR_1 or TR_2. That is, the output is clamped to zero in the absence of a positive trigger pulse to the common base terminal. However, when a positive gate signal is applied between the common base terminal and ground, the base-collector junctions of both transistors are pulse-biased beyond cutoff, thus removing shunting effects of the transistors between the input and output terminals. Consequently, any input signal that is applied during the presence of a positive gating pulse will be directly coupled to the output terminals by way of capacitor C_C.

Before leaving the subject of gates, it may be well to briefly investigate the application of a purely magnetic two-state element which is rapidly becoming prominent in digital computer circuitry. (See also *Saturable Reactors* as treated in the section on *Amplifier Components of Computers.*)

Two-State Magnetic Core Devices, provided with properly polarized input and output windings, have been recently developed for high frequency switching and gating circuitry. Consider, for example the *Two-State*

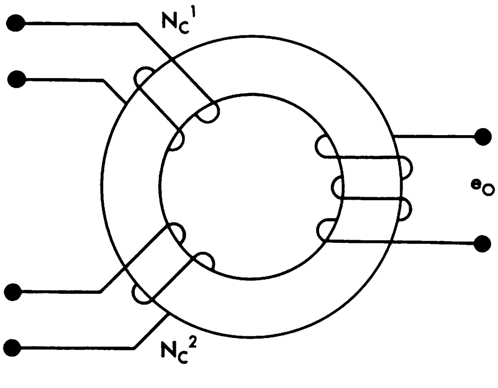
Saturable-Core Reactor and its accompanying square-loop-hysteresis characteristics illustrated.

If a positive pulse of sufficient magnitude is applied to the coil $N_C{}^1$, current flowing through the turns of this coil tend to saturate its common toroidal core with magnetic lines of induction. The ampere turns (*mmf*) produced by this initial positive pulse may take the magnetic state of the core to point *a* as indicated. It must be remembered that cores of this type possess the ability to retain their flux level (in the form of residual magnetism) even after the magnetizing force has been removed. Hence, upon termination of the initial positive pulse applied to $N_C{}^2$, the toroidal flux relaxes to point *b* and remains in its positive remnant state of induction.

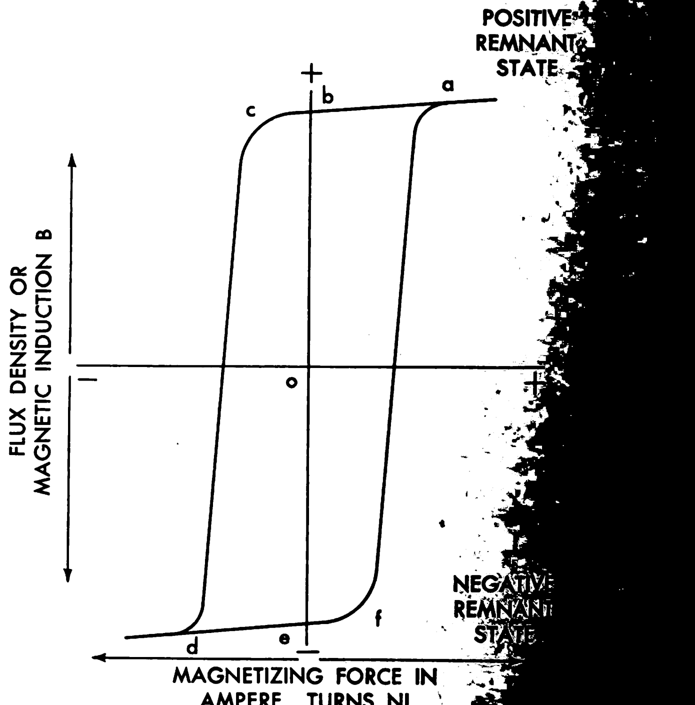
If at some later time a negative pulse is applied to $N_C{}^2$, so as to have an opposite magnetizing effect, the toroidal core undergoes a transition from the positive saturated state, through zero, to the negative saturated state. The flux density then relaxes from point *d* to *e* when the negative excitation pulse is terminated, and the toroid now remains in its negative remnant state until subsequently triggered with a pulse of the original polarity.

During the transition, the toroidal core undergoes a large change in flux from point *c* to *d*, resulting in a correspondingly large relative motion between this flux and the output coil windings. Hence, by normal transformer action, a pulse is produced at the output terminals of the magnetic core whose polarity is determined by the manner in which the output coil is wound.

Next consider one of the input windings to be biased with a DC reference current so as to maintain the toroidal core in its positive remnant state. As a result, the negative pulses, when applied individually to either $N_C{}^1$ or $N_C{}^2$, have insufficient magnitude to demagnetize the core below its positive knee of saturation. Thus, noncoincident negative inputs produce a very small change



TWO-STATE SATURABLE CORE REACTOR



SQUARE LOOP HYSTERISES CHARACTERISTIC

in flux, which results in an insignificant output.

However, if negative pulses are applied simultaneously to both input windings, the polarizing (bias) current is suddenly overcome. This allows the state of the core to rapidly switch over from its positive to negative remnant condition which results in a large change in flux and, in turn, a large induced output pulse.

There are, of course, many variations of magnetic core gating devices. Toroidal cores may also be interconnected in various series-parallel combinations to form delay lines and magnetic-core memories.

A very new solid-state element (even more minute than present day transistors) called the *cryratron* has recently been developed. The cryratron consists of a tiny length of wire with another small wire wound around it in the form of a coil. The cryratron employs a magnetic field to gate the super-cooled wire core *in and out* of superconductivity. Compact liquid helium refrigerating units hold the cryratron elements at several hundred degrees below zero. Under this supercooled condition the minute wire cores are superconducting until influenced by a controlling magnetic field.

Hundreds of these elements may be arranged in a matrix form requiring only a few square inches. Because of their compactness and economy, it is likely that these new elements will lend themselves well in expendable flight control computers utilized in unrecoverable interceptor aircraft or missiles.

COUNTING COMPONENTS. Another important function which takes place in the arithmetic or logical unit is that of counting. In any digital computation, there is a necessity for a means by which the number of operations may be determined. Devices which perform this function are called *counters*. Counting may be performed in connection with a punched tape or magnetic recorder. In other words, the repetition of pulses can be recorded along with other operational functions. Mechanical or electromechanical devices such as multiposition switches can be rotated by an electromechanically operated clutch system to give any desired number of pulses in a given period of time.

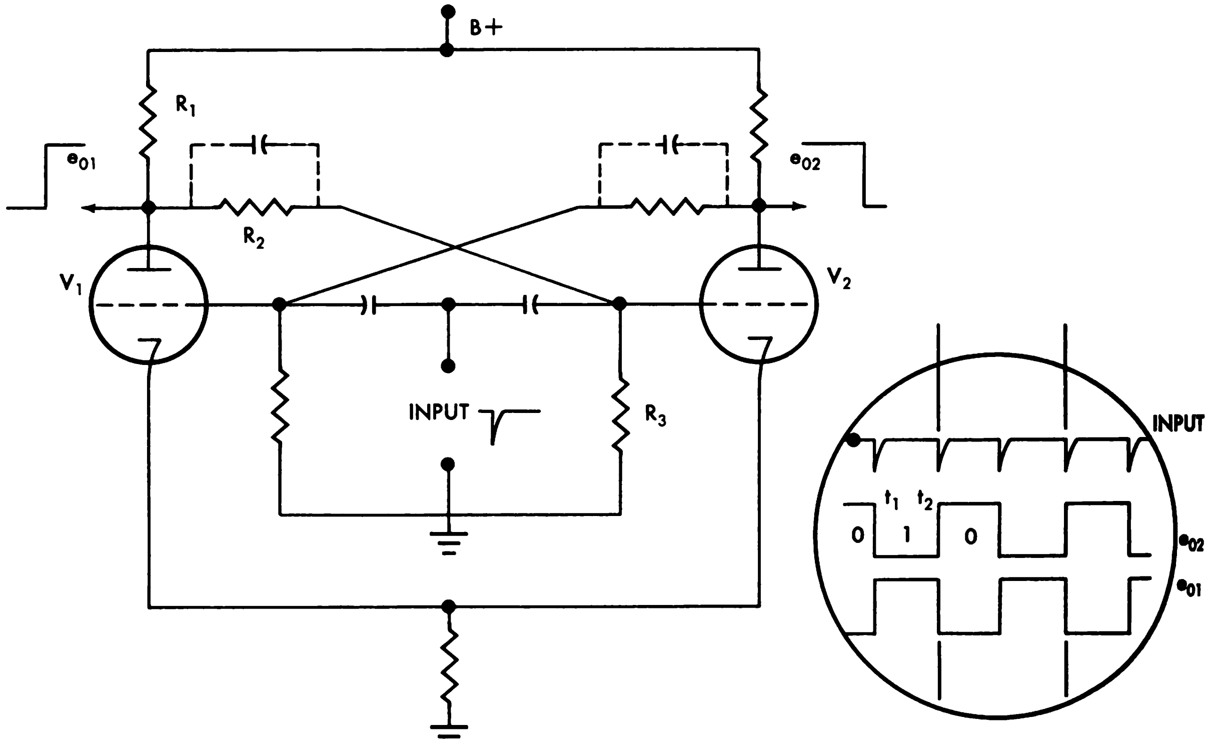
The counting function may or may not be a continuous operation in a computing system. The point in time when a counting function is to take place may depend upon the state in which other components of a system are preset. That is, it may be necessary to count a certain number of other operations that have taken place before some new operation is to begin.

One of the basic methods for providing either a continuous or an interrupted system of counting involves the multivibrator or flip-flop principle. The most common form of the multivibrator uses two triode tubes having plate-to-grid feedback from one tube to the other and a common bias arrangement between the two halves of the circuit.

The *Eccles-Jordan Trigger Circuit* illustrated is the most common type used. It is a special form of multivibrator which employs direct coupling between the plates and grids of the two tubes. Note that the circuit resembles two DC amplifiers connected back to back. Variations of this circuit are used as subharmonic oscillators or as component parts of either decade or binary counters.

The operation of the Eccles-Jordan trigger circuit is not dependent upon a continuously applied driving voltage. It will respond to a single input pulse by reversing its state of equilibrium and will remain in the new state of equilibrium until the next input pulse is received.

A voltage to trigger a circuit of this type may be applied to both grids simultaneously or separately, and it may be of positive or negative polarity. Ordinary noise voltages and minor plate voltage fluctuations will not disturb circuit equilibrium, nor will pulses or inputs having gradually sloping wavefronts. However, a sharp transient will disturb the circuit equilibrium sufficiently to cause the conducting tube to cut off and the nonconducting tube to conduct. Pulses of



Eccles-Jordan Trigger Circuit

sufficiently sharp rise time may vary in duration, frequency, and amplitude; yet the circuit will respond as long as the trigger pulses are sufficiently spaced to permit the circuit to cycle completely.

The waveforms accompanying the diagram illustrate the action when a series of negative trigger pulses is applied after an initially assumed state of equilibrium has been reached. Note that in this particular example the initial stable state is represented by V_1 conducting and V_2 cut off and that the transition from one state to another is brought about by applying a negative pulse to the input terminal which is common to both grids.

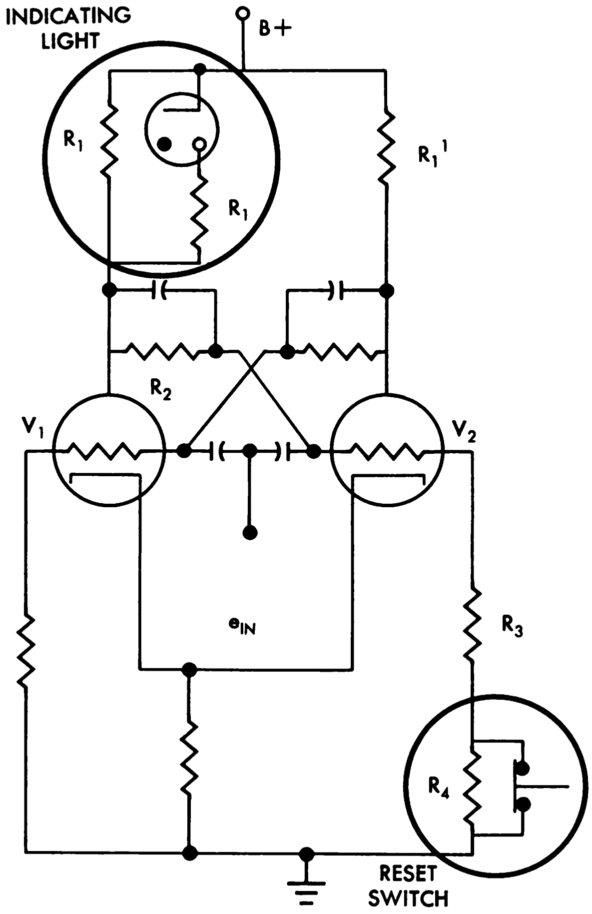
Since V_2 is initially assumed to be in its cutoff state, the injection of a negative trigger pulse to its grid at time t_1 can produce no direct effect on its plate current flow. The same trigger pulse will, however, have a considerable effect on the conduction of V_1 which, in turn, affects the conduction of V_2. That is, the application of the initial negative trigger pulse at the input drives the grid of V_1 negative, causing a decrease in current flow through this tube which permits its plate voltage to rise. This positive rise in V_1 plate voltage is directly coupled to the grid of V_2, driving it into conduction. Since the current flow through V_2 increases, there results a corresponding drop in its plate potential. This now negative-going pulse (plate voltage of V_2) is, in turn, directly coupled (fed back) to the grid of V_1, which augments the negative voltage already there from the initial trigger pulse and drives V_1 below cut off. Hence, a cumulative action takes place, causing the circuit to be switched over rapidly to another stable state with V_1 now cut off and V_2 conducting. The process described above is, of course, reversed when t_2 (the next negative trigger pulse) is applied at the input, and the circuit will therefore be switched back to its original stable state.

From the foregoing discussion, it is evident that the Eccles-Jordan multivibrator executes one alternation for each trigger pulse, two pulses being required to bring about a complete cycle of operation. Hence, you can see that this multivibrator circuit is suitable for a binary system using the digits *0* and *1* —nonconduction of one tube could represent *0*, while its conduction could represent *1*.

Indicating Lamps. The trigger circuit discussed in the previous paragraph makes an excellent count indicator after the addition of several circuit features. The first would be to add some type of indicating device to show that the circuit has been triggered. There are several methods whereby this may be accomplished, but the simplest and most economical is the arrangement shown in the accompanying diagram of *A Modified Flip-Flop Counter*. By placing a small neon bulb across one plate load resistor as shown, for example, a direct visual indication of a tube's conducting state is obtained. If V_1 is conducting, there is sufficient voltage drop across R_1 to ignite the neon bulb. If V_2 is conducting, V_1 will be cut off, and the neon bulb will not light. R_1 is placed in series with the indicator to prevent circuit unbalance when the tube fires. When the tube fires, it effectively shunts R_1. R_1 prevents this shunting effect and is usually larger than R_1. If V_2 is made the initial conducting tube before any trigger pulses are applied, a trigger pulse will cause the neon bulb to light when V_2 cuts off and V_1 conducts, indicating one input pulse.

Reset Switch. It is desirable to have all neon lights extinguished before taking a count to indicate an initial zero reading. It will therefore be necessary, with the indicating light system used, to have all right-hand tubes initially conducting and all left-hand tubes cut off.

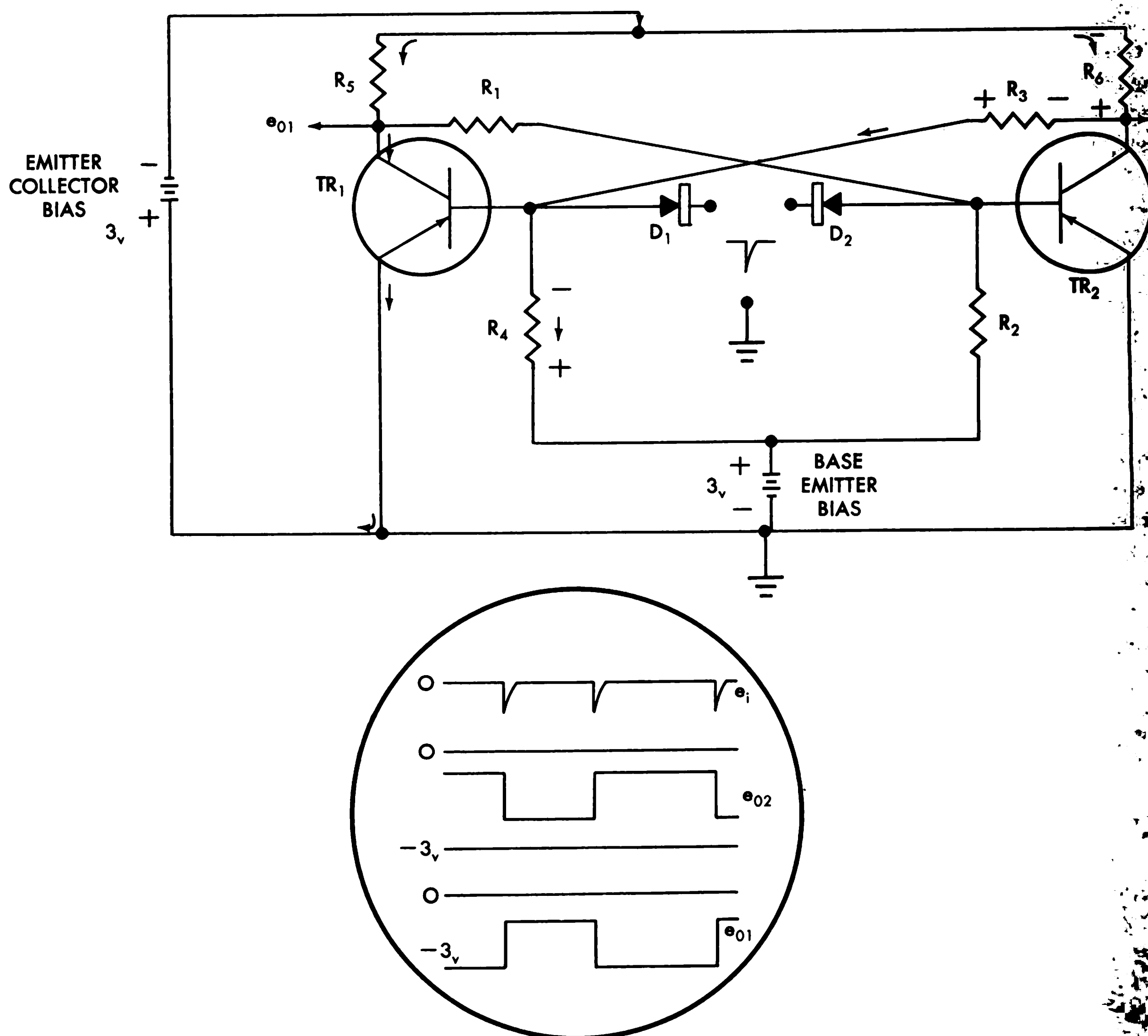
In the modified flip-flop counter, R_3 is shown to be supplemented by another resistor R_4 and a switch which normally shorts out R_4. This switch is a normally closed microswitch (NCMS). With the switch closed, the circuit is balanced. However,



A Modified Flip-Flop Counter

opening the reset switch adds R_4 to R_3. This raises the voltage at the junction of R_2 and R_3, and thus lowers the bias on tube V_2. If R_4 is of sufficient magnitude, V_2 can be made to conduct regardless of V_1 conduction. After V_2 is made to conduct, the switch may again be closed. There are other means of reset; in fact, almost any means to cause either V_1 to cut off and V_2 to conduct may be used. Some of these may be the addition of an external bias voltage to V_1 or the removal of such a voltage from V_2.

A Transistor Flip-Flop Trigger Circuit comparable to its vacuum tube cousin is illustrated next. The initial stable state of this circuit is assumed to be when TR_1 is cut off and TR_2 is conducting. Now let us assume that the transition from one state to the next



A Transistor Flip-Flop Trigger Circuit

is brought about by applying a negative pulse to the base-emitter circuit of TR_1 by way of D_1. The application of a negative pulse to the base of TR_1 increases its forward base-emitter potential, causing it to suddenly conduct. This conduction (of TR_1) reduces the negative potential at its collector; hence, a positive-going pulse is coupled through R_1 to the base-emitter circuit of TR_2, cutting this transistor off. With TR_2 now cut off, its collector potential becomes more negative, approaching the value of the emitter-base bias supply voltage, and the voltage divider, consisting of R_3 and R_4, holds the base of TR_1 sufficiently negative to maintain its conduction even after the termination of negative input trigger pulse. Consequently, the circuit remains in this stable state with TR_1 conducting and TR_2 cut off until the next negative trigger pulse is applied, at which time the circuit will be switched back to its original stable state as shown by the waveforms accompanying the transistor flip-flop diagram.

It should be pointed out that, if the time interval between input pulses is relatively long, it can be assumed that a multivibrator has stored the effect of one pulse until the next pulse is received. When required, the effect of the first pulse is released, and the new effect stored until the arrival of another pulse. This characteristic has led to the application of this type of circuit in the storage or memory unit of computers as well as in their counting units.

Memory or Storage Units

One of the most important units in the makeup of modern data processing computers is that unit which performs the operation of storing digital data. Because storage devices have many features in common with our memory processes, they are frequently referred to as *mechanized memories.*

The storage unit is the section of the computer in which information is retained for short or long periods of time until its use is required in the solving of a particular problem.

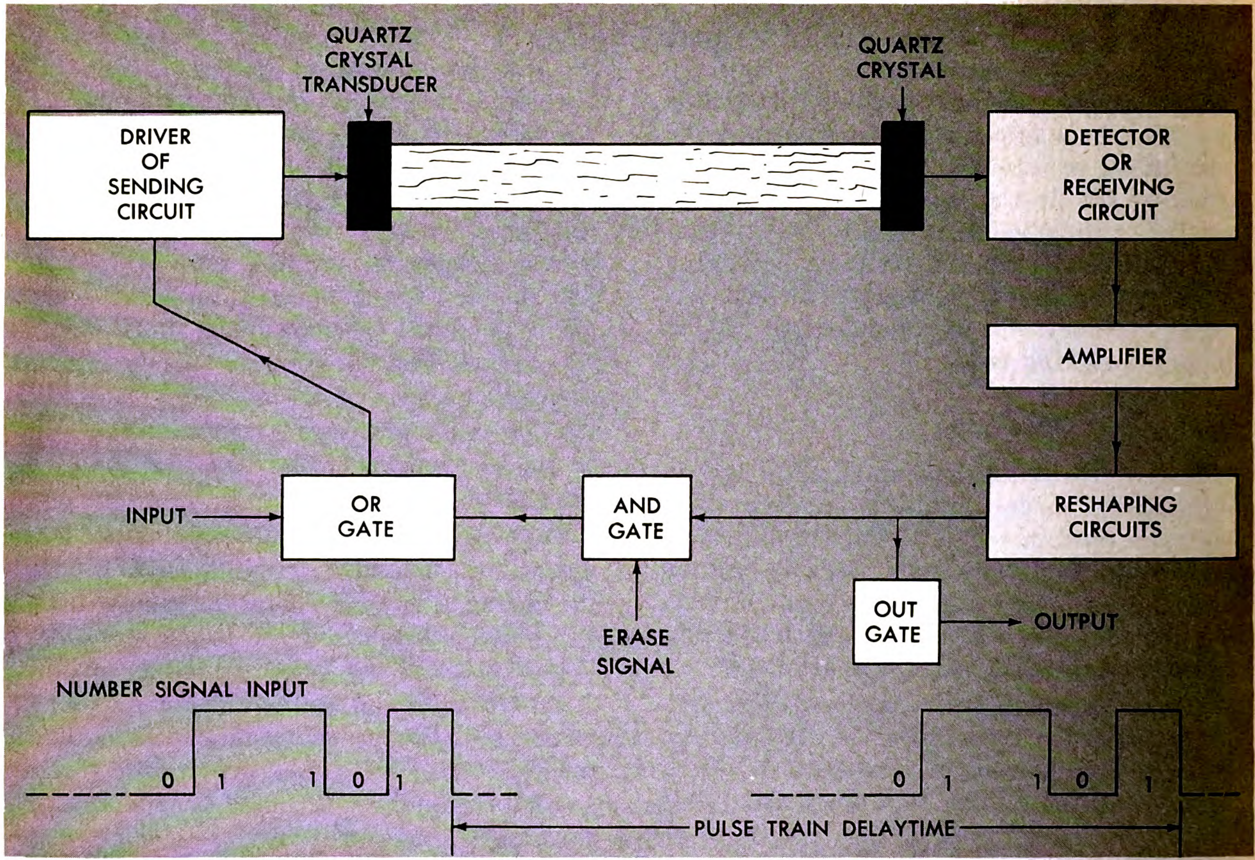
Frequently, you will find the term *delay circuit* used in connection with the storage unit. This delay is actually a storage for a short interval of time. The usage of the term *delay* is generally applied to a particular operation that is to be performed after some other operation has been completed. On the other hand, storage refers specifically to the retention of some information that will be required in later calculations.

The diversity of components used in the computer is especially evident in storage circuits. Some of the components that you found to be common to the input and logic units are found again in the storage circuits. The punched tape, magnetic recorder, switches, relays, and Eccles-Jordan trigger circuits are examples of such components.

ACOUSTIC OR SONIC DELAY-LINE STORAGE. The *Mercury Tank Sonic Delay Line,* illustrated, is also referred to as a sonic-circulating memory which stores information in the form of a train of pulses or wave patterns. As its name implies, it operates by transmitting pulses at sound frequencies through a medium of liquid mercury. Notice that the mercury is contained in a steel tube capped at either end with a quartz crystal. One crystal serves as a receiving transducer, and the other as a transmitting transducer. Quartz is especially suitable for this application due to its piezoelectric qualities (electrical potential developed due to pressure on a crystalline substance). Quartz is also desirable because of the fact that its acoustical impedances insure a maximum transfer of sound energy between the mercury and the quartz.

Suppose that we feed into the input of this delay line a message or number signal with the coding *10110.* The quartz crystal will receive this coded pulse train (*pulse—no pulse —pulse—pulse—no pulse*) by way of the *input-OR* gate and will set up sound waves within the mercury that have the same digit pattern. Thus, trains of sound waves representing the digits of the number signal travel through the mercury slowly as compared to the velocity of electrons in a current carrying conductor. The amount of delay, of course, depends upon the velocity of sound through the medium and the distance between the quartz transducers. The sound waves upon striking the quartz crystal at the opposite end of the delay line set it into vibration at the same digit repetition rate as that of the quartz sending element. Hence, the input signal (train of pulses) is reproduced in the form of minute piezoelectric pulses. These pulses of energy are then electrically detected, amplified, reshaped, and returned through the *OR* gate back to the input of the delay line where the cycle of energy transfer may be repeated. This cycle of transmission and reception of impulses through the delay line may continue until such time that a gating circuit operates to either transfer the number signal to another circuit or clear it from the system.

The *AND* gate prevents the delay-line output from reentering the input side of the



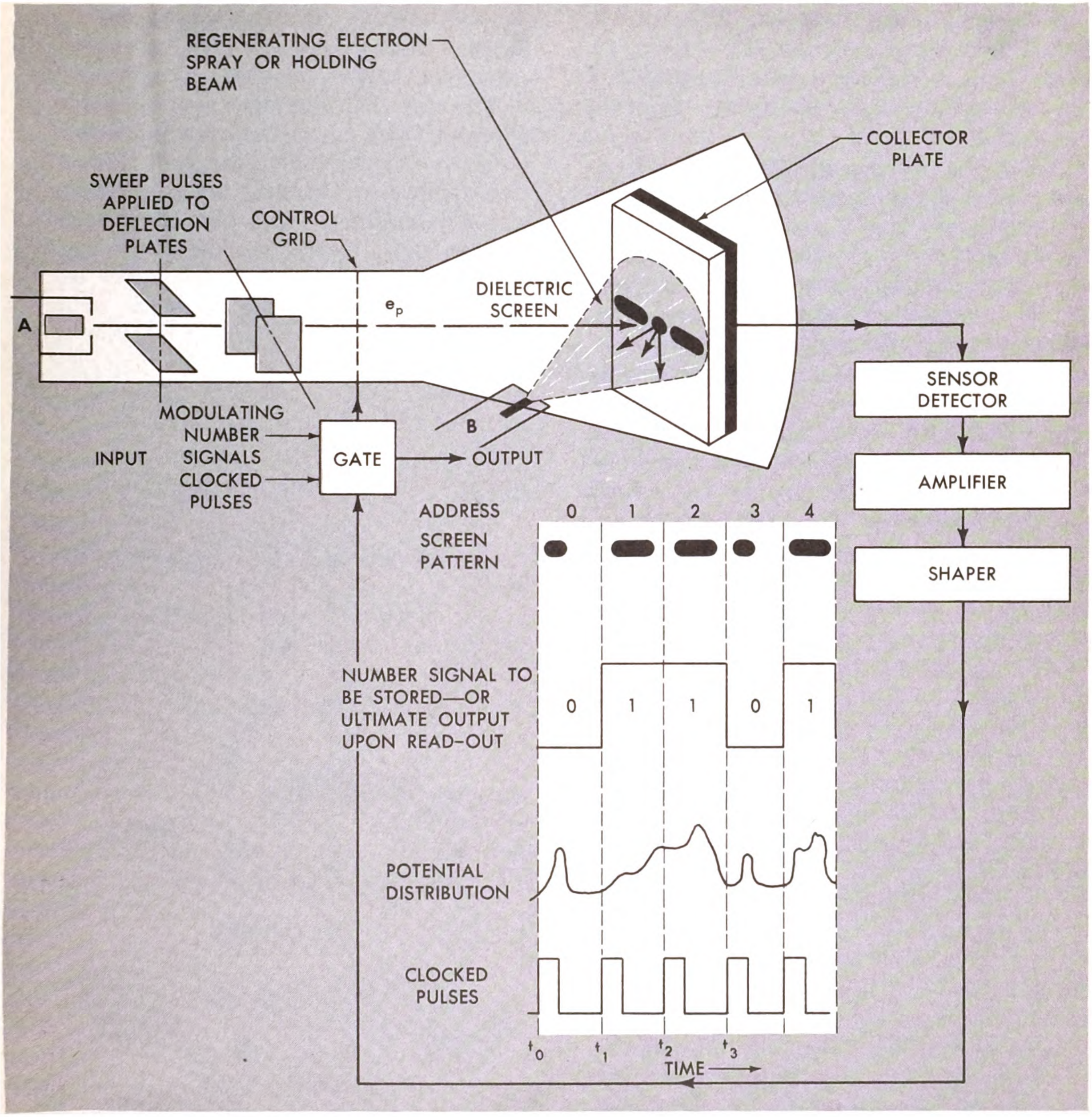
Mercury Tank Sonic Delay Line

system when an erase or clear signal is applied. Such would be the case if a new number signal were to be applied to the *input-OR* circuit, thus allowing the new message to replace the original one in the delay line.

An adding circuit may be incorporated into the circuit in place of the *OR* gate. Under this condition, if a new number signal is applied to the input in the absence of a erase signal, the new number will be added to the number stored, and their sum will be then introduced into the line.

Delay lines of this type are capable of storing trains of several hundred pulses. However, since velocity of sound propagation is sharply dependent upon the temperature of the mercury, the unit must be maintained at a rather constant temperature. Hence, mercury delay lines are usually enclosed in a thermostatically controlled container. The mercury-memory container may have either one or many delay lines.

Electrostatic CRT Memory. An example of an electrostatic storage device, *Cathode Ray Tube Storage,* is illustrated. In this example, a primary electron beam (e_p) from cathode *A* is directed upon a surface (screen) composed of alternate conducting and dielectric layers. When this electron beam is pulse-intensity modulated, the screen acquires a train of small localized electrostatic charges. The magnitude of these localized charges depends upon the instantaneous intensity of the *A beam,* the potential of the collector surface, and the secondary emission characteristics of the bombarded surface. If the primary beam of incident electrons is turned on at time t_o to a spot corresponding to the digit *0* of a number signal and then immediately cut off, the magni-



Cathode Ray Tube Storage

tude and duration of the potential distribution are both small. This is shown in the accompanying waveforms at the zero address. The term *address* is a name or number location identifying a register or cell where digital information may be stored—a particular localized spot on the dielectric screen in this example. Note that the resultant potential distribution shown at the zero address produces an image in the form of a dot.

If, on the other hand, the primary electron beam is next turned on at time t_2 to the spot corresponding to the digit *1* and left on while the beam is displaced by the horizontal sweep, the resulting potential distribution is

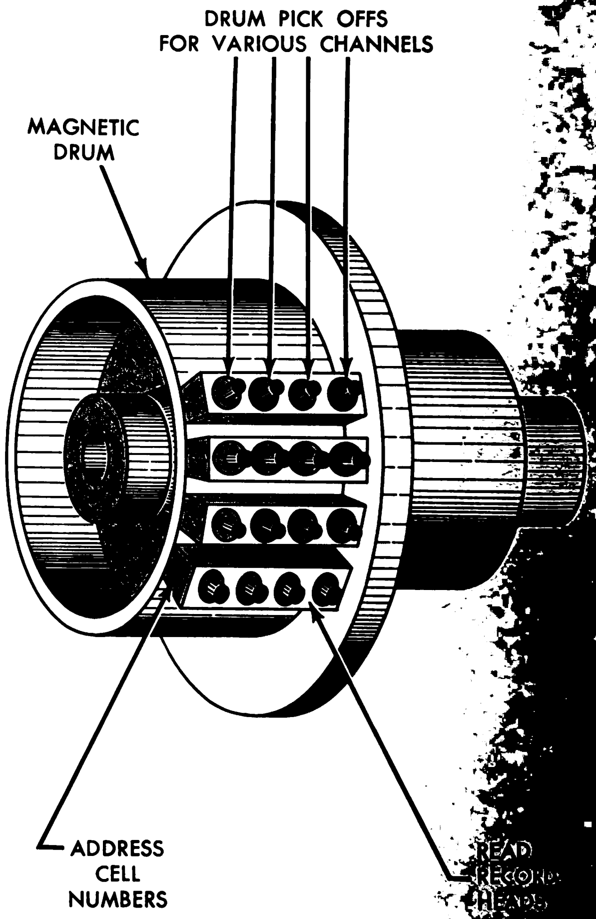
both larger in magnitude and in time duration. This produces an image in the form of a longer dash, which represents the digit *1* of the number signal as shown at addresses *1*, *2*, and *4*.

Hence, the presence or absence of a localized charge at a particular spot on the screen may represent the *1* and *0* digits in the binary system. In this manner, number signal information may be stored on the screen as the presence or absence of these localized electrostatic charges.

More than a thousand units or bits of information may be stored in a CRT of this type, and with some screen coatings, a charge once placed upon a certain spot remains there, unaided, for as long as five minutes. However, the concentrated or bunched electrons producing the localized charge on a small part of the screen have a tendency to fade or be erased during the process of read-out. The read-out function consists of sensing, reading, and copying or transferring a word, such as a number signal, from one form of storage to another—perhaps from an internal storage unit to an external output unit. As opposed to clearing a storage unit, it is often desirable to hold or retain digital data contained in one storage unit even after transferring the message into another storage unit. Regenerating or intensifying the localized charges with a holding-beam will restore the electrostatic charged spots on the dielectric surface. The holding-beam, as shown in the illustration, consists of a diffused beam or spray of electrons emanating from cathode *B*, which may or may not be gated. Read-out may be accomplished upon the application of clocked-gate pulses in coincidence with the potential distribution pulses which are present because of the localized electrostatic charges. That is, output pulses proportional to the potential distribution pulses are detected at the collector plate, amplified, reshaped, and gated to an external storage or output unit.

Magnetic Drum Memory. Number-signal data may also be stored as a pattern or train of *magnetically* polarized spots upon surfaces of materials which have high magnetic remanence characteristics. Because of the closeness with which data may be stored on these surfaces, magnetic tapes and wires are used considerably in input and output equipment. However, magnetic tapes or wires are not suitable for use in a high speed internal storage unit. Hence, the difficulties associated with handling tapes and wires led to the development of the magnetic drum, where the magnetic recording medium (perhaps a thin coating of magnetic iron oxide) is applied to the surface of a nonmagnetic cylinder which is designed to rotate.

At the present time, the magnetic drum memory is considered to be the most compact



Typical Multiple Channel Magnetic Drum Storage Unit

of all the magnetic systems. Because of the closeness with which magnetically polarized spots can be assembled on the surface, the drum itself may be physically quite small—perhaps only a few inches long and few inches in diameter—and yet allow a large number of digits to be written and stored with a high degree of permanence.

Multiple Channel Drums may be used as a *parallel-access* store having several input-output elements for each channel or as a *serial-access* store having one input-output arrangement which may be switched from channel to channel as directed.

The sketch illustrates a *Typical Magnetic Drum Storage Unit* consisting of several recording channels, side by side, with one read-record head per channel.

The general operation of magnetic drum storage units of this type may be more readily understood by considering only a single cell within a particular channel and only one read-record head.

A *Single Channel* or section of a typical magnetic drum storage unit is illustrated next in the functional diagram. Only one read-record head is shown to be either recording (writing innformation into) or reading (getting information out of) the various cells contained in the single channel as the drum revolves.

The uppermost waveform, accompanying the functional diagram, represents a pulse-frequency modulated message to be stored—perhaps for later use. In this waveform the digit *1* is represented by a frequency lower than the center (resting) frequency, and the digit *0* is represented by a frequency higher than the center frequency. This coded PFM signal is detected within the receiver and converted into the pulse-no-pulse wave pattern shown directly below. This number signal is next gated through the writing portion of the read-record head.

The drum shown is assumed to be revolving at a constant speed; however, the *pulse-no-pulse digits* of the number signal will present themselves to the various localized cells on the surface of the drum at a frequency which is made to coincide with the clock pulses controlling the computer. Hence, a magnetic pattern—a sequence of locally magnetized cells or spots—is set up upon the surface of the rotating drum.

Consider, for example, that while the magnetic drum is in the position indicated, the first positive (shaded) pulse representing the digit *1* is gated to the writing circuit. A pulse of current is passed through the writing coil encircling the magnetic core, and a magnetic field is built up. A portion of this field spreads to the air in the vicinity of the air gap and, in fact, expands so as to magnetize the localized portion of the drum surface indicated as *cell 4*.

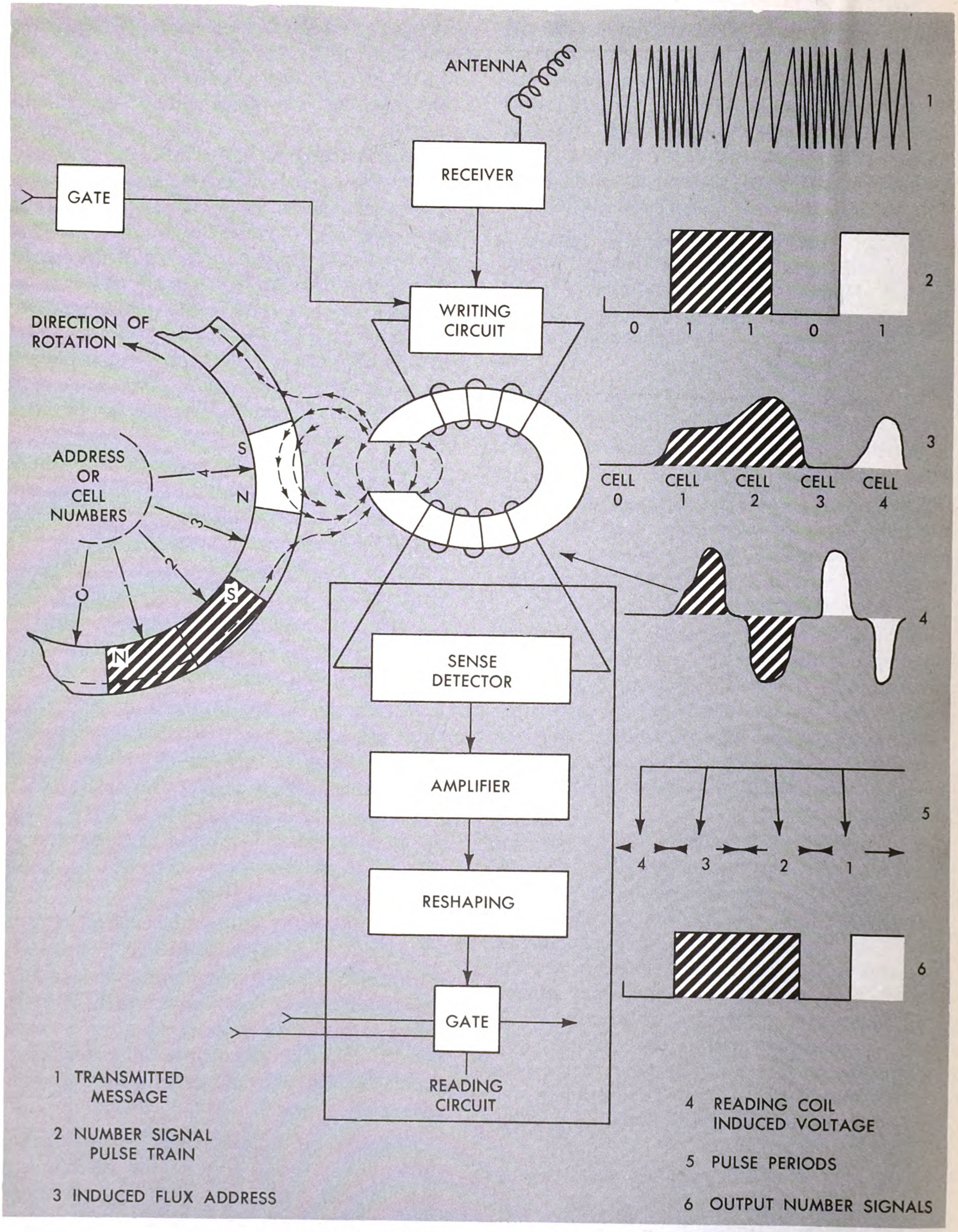
The polarity of this magnetized spot and the induced flux corresponds to the direction in which current flows through the writing coil of the read-record head.

During the next pulse period—the second pulse period representing the digit *0*—cell number *3* is unaffected, because in this example a permanent magnetic bias is assumed, representing all zeros.

During the third and fourth pulse periods, two consecutive positive pulses—representing the digits *11*—are to be stored. Hence, the leakage flux that is to be subsequently produced by this portion of the number signal will persist for a longer time duration—two pulse periods—and thereby magnetize *cells 2* and *1* as indicated.

Thus, digits of the number signal may be stored as localized magnetic spots upon the surface of the rotating drum. With a digit spacing of 100 per inch on a 4-inch diameter drum rotating at 3600 rpm—with a corresponding pulse period in microseconds—thousands of bits of information may be stored per channel.

Read-Out is accomplished when clocked pulses are directed to gate the reading circuit and the individual cells present themselves to the read-record head at the proper frequency. Hence, the digits contained in the



Single Channel Functional Diagram of a Typical Magnetic Drum Unit

number signal will be read-out within one revolution of the rotating magnetic drum. That is, as the localized magnetic spots (cells) pass the air gap of the read-record head, flux will vary in the reading coil. The relative motion between this varying flux and the turns of the reading coil causes a voltage to be induced into its windings. The form of this voltage is shown to be slightly larger—both in magnitude and in time duration—when consecutive magnetized spots (*cells 2* and *1*) are presented to the read-record head.

This pattern of induced pulses is then sensed, detected, amplified, reshaped, and gated to some external unit such as a register or arithmetic unit.

Switches and Control Unit

Having looked at many of the individual bits of circuitry contained in the input, logic, and memory units, it is now possible to deal with the general organization of the switching and control circuits and, at the same time, review the general sequence of digital computer operation.

First, the coded program of instructions is decoded and stored by the input unit. Once the computation or processing of the data is initiated, the programmed information is fed at a clocked pulse rate into the main memory or storage unit. Initial instructions are then fed to the switching and control unit, where they are stored temporarily until carried out. The control unit then instructs its associated equipment when to perform a sequence of operations. The instruction may call for the transfer of data from the main memory unit to the arithmetic unit for temporary storage. Information, in the form of a number signal, may be temporarily stored in the arithmetic unit until further number signals are received, at which time the combination is operated upon and the result again fed to the main store. Upon completion of all instructions the result of the computation is transferred to the output unit where it is usually stored in its final form.

COMPOSITION. The control unit, then, actually consists of many counters and logic switching circuits which gate the instructions from the main storage unit in the proper sequence and transmit the required command signals to other associated computer circuitry. You will find that the control unit contains an *instruction counter* and *instruction register* as well as *function* and *address decoders*.

Instruction Counters and Registers consist of one or a series of binary cells or elements (flip-flop or magnetic core type) which are capable of storing digital information such as an ordered set of characters or digits like the number signal *1101*.

Function and Address Decoders consist of a circuit or a network of circuits having a number of inputs and outputs that are interconnected so as to enable coded signals to be decoded—that is, when coded signals are produced which are a function of the input information. You may recall that the term *address* is simply a character or number identifying a register or cell where information is stored.

Number signals as well as address information and instructions may be stored in various coded forms.

If the information is in binary form, the instructions may be represented by the first two or more digits, depending upon the complexity of the operation and the coding system employed. For example, the following table illustrates a typical instruction-operation code:

Instruction	Operation Code
stop	00
add	01
subtract	10
multiply	11
clear and add	100
store	101
transfer	110
transfer or minus	111

The circuit receiving a coded instruction is placed in the proper operating condition in order to perform the desired function upon receipt of the number signal. In some computers used in command guidance systems, a circuit may remain in the operating condition dictated by the last command signal until it receives the next command. In other computers, a circuit may perform the operation as instructed and then automatically return to a neutral or inoperative condition. It is desirable that instructions be reduced to a simple form to eliminate the need for complex control circuitry and to permit a greater portion of the computer's operating time to be used in processing the input data. Although the control data may be fed into the computer by any of the input methods previously mentioned, it is general practice to employ the same input method for both control and information data.

Output Units

After computation processes have been completed, there must be some means for presenting the results to an operator or some system under the control of a computer. Any of the devices that were used for putting information into the computer may also be used to take information from the machine. These include tape recorders of all types, oscillographs and oscilloscopes, and relays with indicator lamps to register their positions. Geared indicators similar to those used in the common desk-calculator may also be used to register the results.

In an airborne interceptor system, the results of the computer may be used directly to influence the control system or perhaps to influence the sensor unit in a directional guidance system. Data from piloted or pilotless interceptor aircraft may be collected and recorded by photographic means or by recording meters with time reference roll charts. This direct recording method is the simplest and most accurate system for obtaining flight data, but is practical only when the interceptor or recording equipment is recoverable. In recoverable airborne equipment, direct-reading instruments using tapes for graphic or magnetic recordings are used as auxiliaries to the telemetering equipment. Photographic systems are used in a similar manner. Thus, the data recorded on tape or film may be compared with that received from the telemetering system, and both the characteristics of the interceptor and the accuracy of the telemetering system may be checked simultaneously.

For some types of data which may be subject to very rapid changes, a conversion link is required between the data source and the recording instrument. Frequency dividers or *down-counters* may be used as intermediate links for slowing down data to some definitely proportionate rate which is within the range of the recording instrument.

ESTERLINE-ANGUS OPERATION RECORDER. One widely used type of recorder, suitable for use either in the interceptor or at the ground control station, is the Esterline-Angus operation recorder. This type output unit consists essentially of four main components: case, chart drive mechanism, writing system, and an electromagnet assembly. Electrically actuated, the instrument may have from five to twenty pens recording the outputs of a like number of data pickups simultaneously. Data is taken from sources where the information is of the *on-off* or *yes-no* type which indicates the time and duration of the data event and how many operations or changes occur. Charts are rectilinear, which makes it easy to compare the recordings with respect to time.

The record charts are of special grade paper and may be obtained with or without time calibration marks along the edges. The chart drive mechanism may be a self-contained spring clock, a self-contained synchronous clock motor, an external motor or timing device, or a combination of these devices. Where time intervals must be meas-

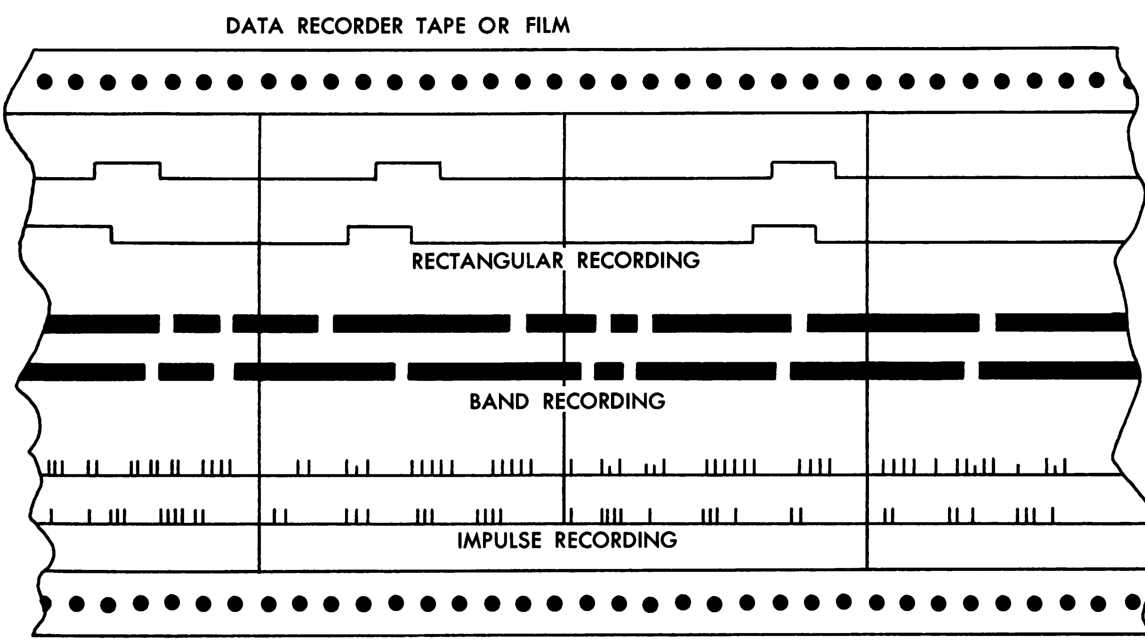
ured accurately to less than one second, a chart speed of at least three-fourths of an inch per second is used. Speeds up to three inches per second are available with the external motor drive. The chart normally travels at a rate of three-fourths inch, one and one-half inches, or three inches per hour. When a data change occurs, the chart is automatically engaged with an external motor by a solenoid clutch and is immediately accelerated to 3,600 times its normal speed. This high speed permits accurate recording of rapid data such as the conducting states of gates and flip-flop circuitry.

Three Types of Record Patterns may be obtained on the chart. These patterns depend upon the relationship between the length of time the pen is energized, the time the pen is deenergized, and the speed of the chart. The types illustrated are known as *rectangle, band,* and *impulse* patterns, so named because of their appearance on the chart.

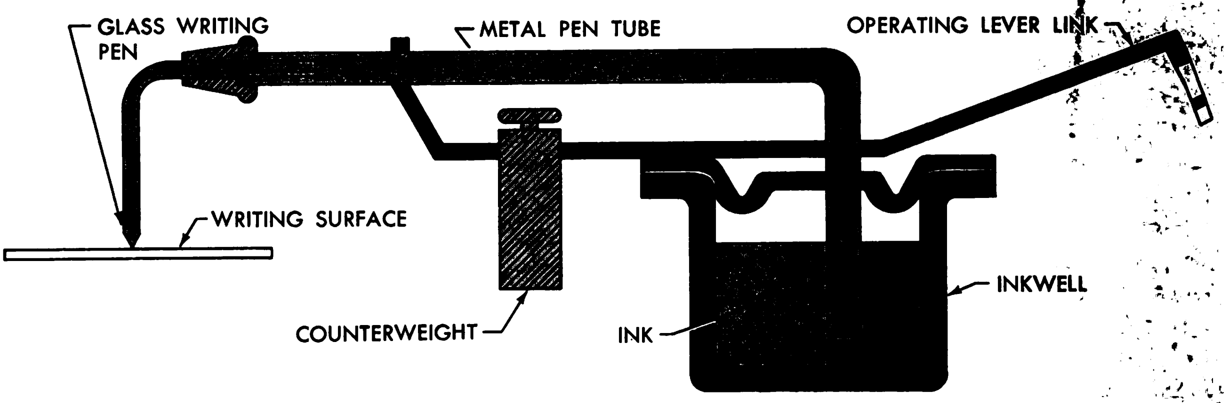
In the Esterline-Angus operation recorder, all of the pens are supplied from a single enclosed inkwell which is designed to mechanically support and align the pens.

A Typical Recording Pen is shown in the next diagram. Note how the pen tube dips into the ink supply through holes in the metal cover of the ink well. Small counterweights fastened to the pen tubes hold the pens on the chart paper at all times. The writing pens are made of transparent glass, providing visual inspection of ink feeding. Divided inkwells may be used so that one group of pens may record in red, another in green, and so forth.

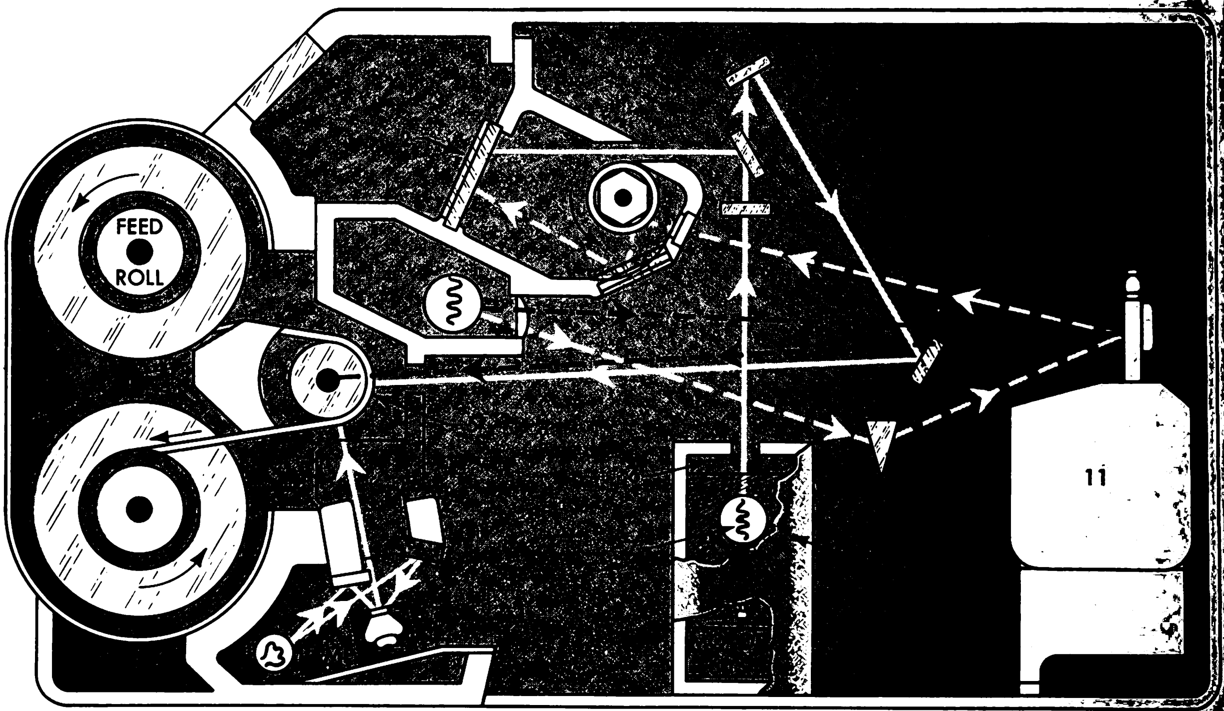
Pressure of the pen on the chart is very light. The driving mechanism has ample power to overcome any drag on the chart which might produce a time error. Each pen is actuated by an electromagnet, causing the pen to move laterally across the chart. Each electromagnetic element will follow as many as ten complete on-off cycles per second providing that the on and off periods are substantially equal. There must be at least .05 second of energized time to allow a pen to



Types of Record Patterns



Typical Recording Pen



1 PHOTO-SENSITIVE PAPER TAPE

2 VIEWING SCREEN

3 ROTATING POLYGON

4 RECORDING AND SCANNING LAMP FILAMENT

5 RECORD DRIVE ROLLER

6 1 SLOT SCANNER DISK

7 TIMING LAMP

8 RECORD NUMBER COUNTER

9 TIMING UNIT

10 GALVANOMETER

11 GALVANOMETER MAGNET

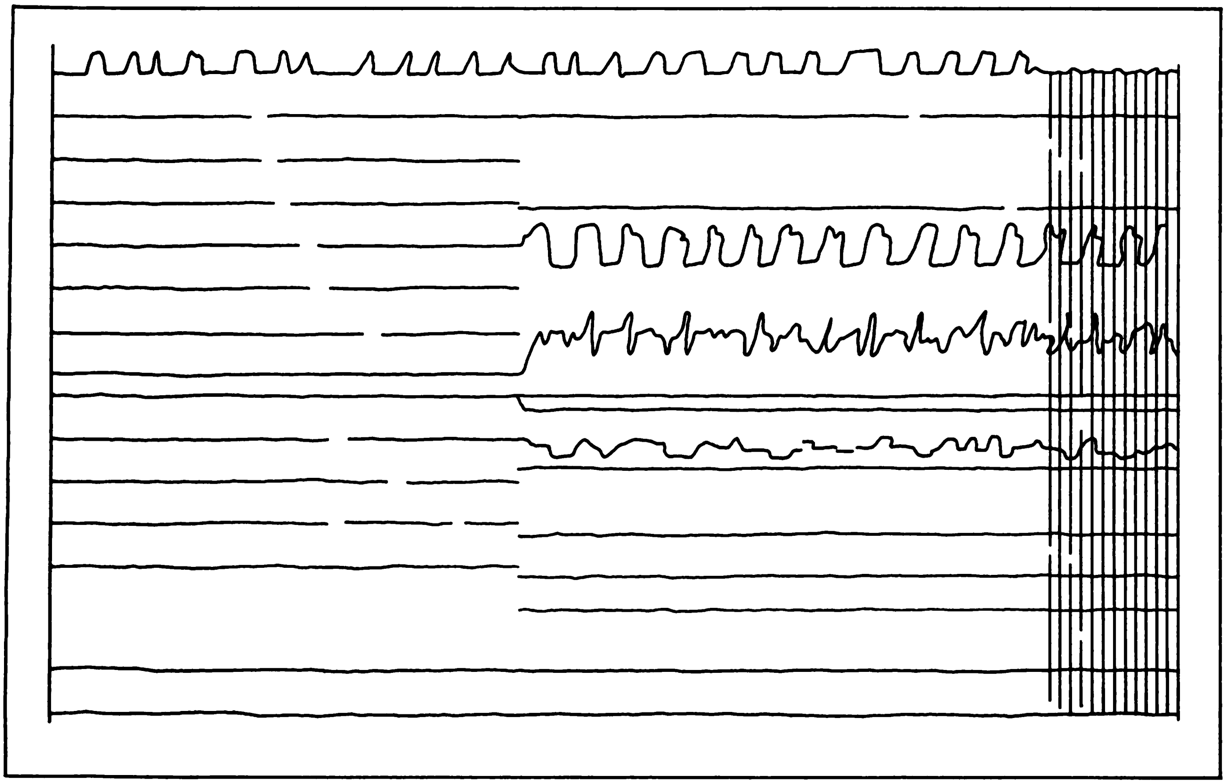
Consolidated Engineering Recording Oscillograph

complete its stroke reliably. If a pen is energized normally, the deenergizing impulse likewise must last for at least .05 second.

For data measurements involving changes of more than ten cycles per second or very wide ranges of voltage variation, devices such as recording oscillographs must be employed.

RECORDING OSCILLOGRAPH. One widely recording device of this type is the *Consolidated Engineering Recording Oscillograph* (illustrated) which contains up to 18 recording galvanometers. These galvanometers deflect a beam of light across a photosensitive paper in accordance with variations of output current in the data pickups or in the outputs of data channel discriminators of a receiver. The photosensitive paper is rolled past the light beam at a constant rate, producing a permanent record on the paper. This oscillograph is a very flexible instrument and is well suited to record telemetering functions. The time base for the recording is the speed at which the paper is run through the recorder. The paper speed is continuously variable over a wide range, and a large selection of galvanometers is available, permitting the recording of many types of data. The machine has been modified so that the timing-line motor is driven from a very accurate 60-cycle per second source. You will find an actual sample of an oscillograph record accompanying the illustration. This is what the operator at the control station sees on the recorder as the event is taking place.

A supplementary recording device such as a magnetic-tape recorder is used in conjunction with the recording oscillograph. The outputs or ground based receivers and calibration oscillators are recorded directly upon the tape. This process is known as *raw data recording* and provides a permanent record



Consolidated Oscillograph Record

of all data received in the event of failure of the oscillograph. If the original oscillograph recordings are defective or if a different time base should be used, the raw data may be played back through the oscillograph at any different rate desired. Frequently, a large number of cross-over galvanometer markings occur on the original oscillograph recording; by repositioning the galvanometers and playing the tape-recorded data back through the system, the confusion of traces can be eliminated. It is also possible to play back selected data channels only, removing all others from the record. The playback records are sufficiently accurate to permit making the data reduction of the resulting record with greater ease than from a single original oscillograph record.

CHAPTER 5

INTERCEPTOR GUIDANCE AND SPACE STABILIZATION

At the present, Air Defense Command is placing its early warning systems farther out from the United States and farther up into the sky. In North America, a giant line of lonely outposts is already probing the skies with unblinking eyes, ready to give instant alert to our airborne and ground control stations. *Distant Early Warning (DEW)* lines stretch across 3,000 miles of the frozen north land to warn us of approaching enemy intruders. Each DEW line outpost is linked with others in the chain through single-sideband carrier equipment which channelizes many types of information for radio transmission.

We also have a line of *Texas Tower* radar sentinels now guarding our shores. Data from these radar islands, along with voice and intercept commands, are *multiplexed* by single-sideband equipment and transmitted on a microwave radio beam. If an intruder approaches, his range, bearing, speed, and course are instantly transmitted to our Air Defense Command.

In the event of an attack, these and other means of detection would be pouring intelligence information into a central point at a rate far beyond human capabilities to digest and act upon. Hence, to meet this problem, our government has developed what the Air Force calls *SAGE,* a Semi-Automatic Ground Environment system. By electronic means, SAGE combines data from many points and with the aid of digital computers solves air defense problems at lightning speed. SAGE may also assist the air defense commander with radarscope presentation of an air battle within a particular geographical area.

The Air Force can use interceptors or tactical fighters, as well as missiles, to destroy enemy aircraft. Within a combat zone, continuous radar surveillance permits control of air space. Interception and destruction may be accomplished through controlled application of surface-to-air missiles (SAM), fighter interceptors, as well as air-to-air missiles (AAM)—launched from jet fighter interceptors—against the intruders.

A recent exhibition revealed the F-101B (*Voodoo*), a supersonic fighter, which can be armed with *Genie* AAM atomic rockets and/or with Falcon guided missiles. In the event of an enemy air attack, a Voodoo-Genie combination may get the first pass at enemy intruders because of the interceptor's long range and vertical climb capabilities.

If enemy craft penetrate our outer defenses and outrun or evade our fighter interceptors, supersonic target-seeking, surface-to-air missiles may be hoisted into firing position and made ready for launching in a moment's notice.

Presently many of our major cities are guarded by *Nike* stations. These station sites

have been chosen according to a master plan to assure the most complete protection. Once launched, Nike is controlled by ground radars. One radar tracks the intruding target, and a second radar tracks the Nike P/I (pilotless interceptor).

Nike control stations utilize one of the later variations of command radar guidance techniques.

COMPONENTS OF COMMAND GUIDANCE SYSTEMS

A command guidance system is one in which the flight path of a P/I (pilotless interceptor) is monitored relative to a target by some agency outside the airframe. From the above definition we find that certain basic operations are of prime importance. First, some tracking agency must record the positions of both the P/I and the target during a mission, either continuously or at designated intervals, or the mission may require that only the P/I be tracked, and not the target. The second operation which must be performed involves interpreting the tracking information and computing the guidance signals which will be transmitted to the P/I. Thirdly, some method must be provided for transmitting the command signals to the P/I. Generally the commands are relayed by some form of radio link. Finally, the P/I must be equipped to receive, interpret, and use the command signals so that they will produce the desired change in the flight path —to correspond to the evasive action taken by the target.

Command Radar Guidance

One variation of command radar guidance employs the use of radar equipment to perform both the tracking and directing operations. This method is especially applicable to SAM, such as the Nike, where both the P/I and the target must be tracked in order to obtain command signals to guide the P/I.

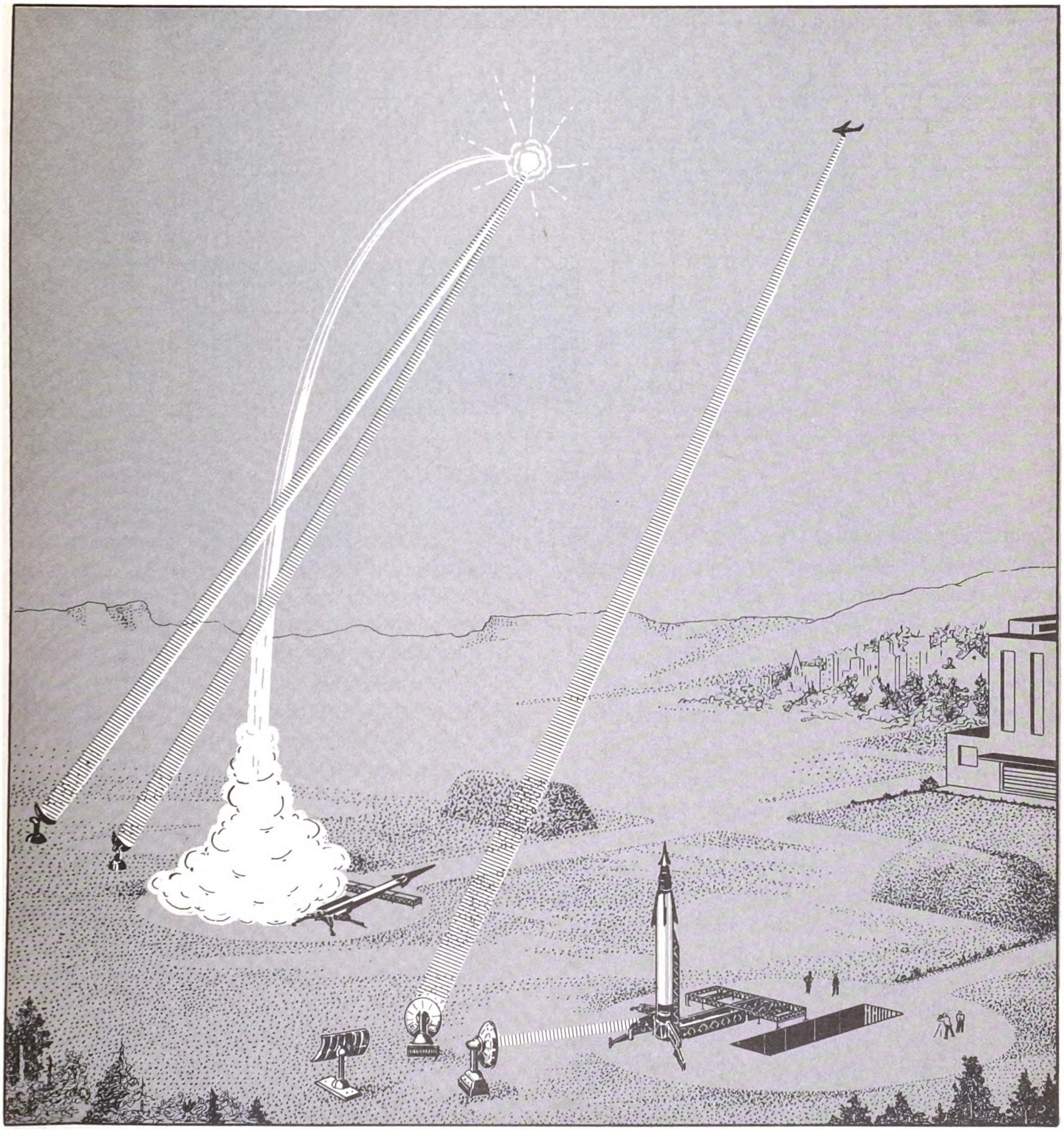
As indicated by the various radar beams in the illustration of *A Command Radar Guidance System,* the target tracking radar obtains a continuous flow of information which gives the slant range, elevation angle, and azimuth angle of the target. At the same time the P/I tracking radar is keeping a continuous account of the slant range, elevation angle, and azimuth angle of the P/I. The information obtained by the two tracking radars is fed to the computer unit which compares the data and issues orders to the P/I tracking-command radar. Thus, the computer controls changes in the P/I flight path so that when the slant range of the P/I is equal to that of the target the P/I will also have the same elevation angle and azimuth angle. When these conditions are satisfied a collision of the target and P/I will result. Usually the P/I uses some type of proximity fuze that will detonate the warhead whenever it and the target come within a prescribed distance.

Command Links

The equipment used to convey command signals to a P/I actually performs the function of a communications link. If a human pilot were in the P/I, voice radio would be used. However, in the case of P/I, the commands must be sent in the language of interceptor guidance and control equipment.

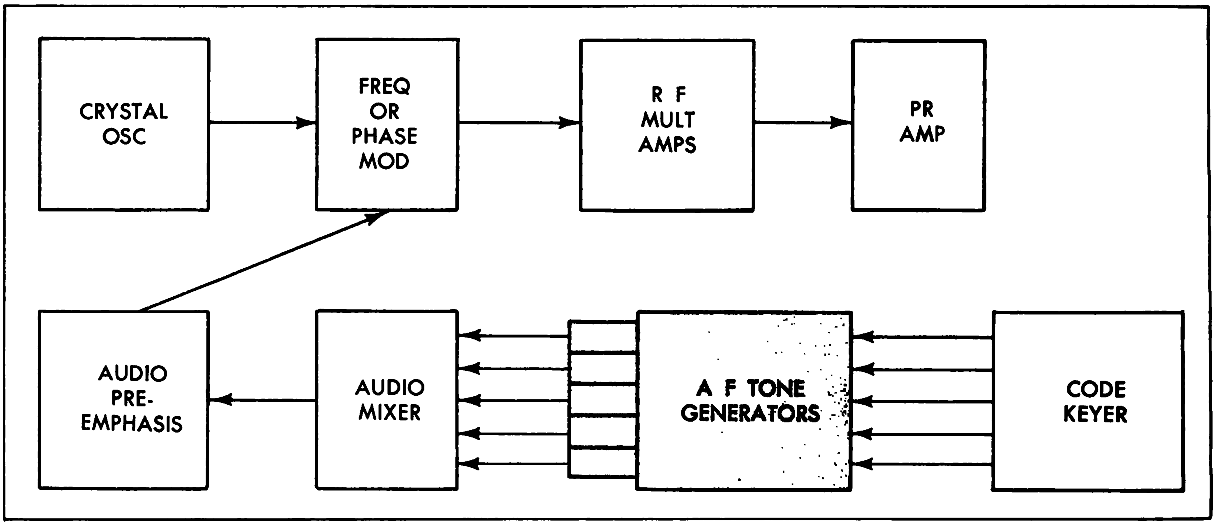
The methods of coding and decoding (modulating and demodulating) RF signals will be investigated in the next section of this chapter. It may be well to first present the general characteristics of transmitter and receiver equipment utilized in command guidance of P/I.

COMMAND TRANSMITTERS. The RF channel of a command transmitter is similar to any other FM or PM transmitter. As shown in the diagram, a quartz crystal is used for stability and accuracy of the carrier frequency. After modulation, there are the necessary stages of frequency multiplication to raise the carrier to frequency in the VHF band. The power amplifier furnishes the RF power necessary to send coded information through space to the P/I receiver.



A Command Radar Guidance System

The peculiarity of the command transmitter is evident in the lower audio channel. Note that a group of *audio tone generators* is present. Each generator operates only when actuated by the circuits in the *keyer* which may be manually or computer controlled. Here coding combinations of the tone channels are devised. Certain command operations of the P/I cannot take place unless several selected tones appear at its receiver



Command Transmitter

simultaneously. Hence, the possibility of an interfering source duplicating this predetermined combination is negligible.

The individual outputs of the tone generators are mixed together into a composite audio signal and applied through the audio pre-emphasis network to the modulator.

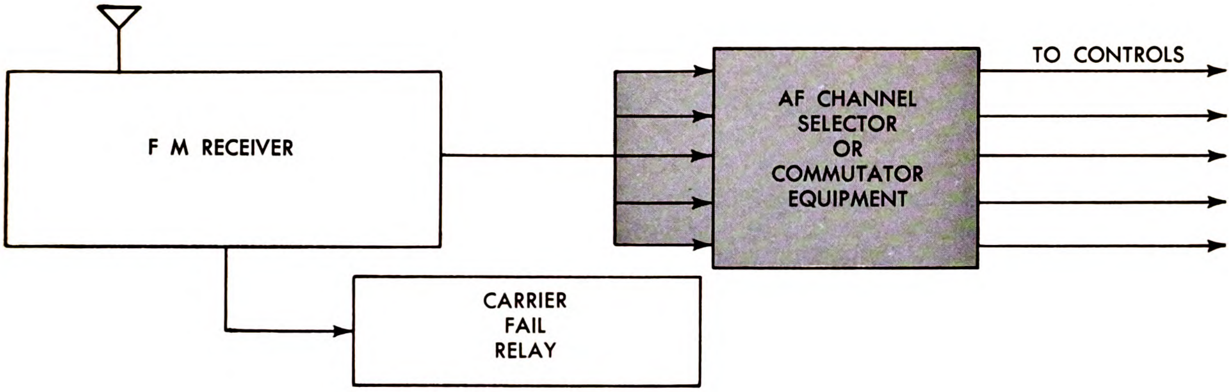
The Pre-emphasis Network is included in order to maintain the audio signal-to-noise ratio at optimum for the overall system, including the receiver. The function of this network is to emphasize the higher audio frequency tones which are later de-emphasized in a network in the receiver having opposite characteristics. This operation causes the signal-to-noise ratio to remain more constant throughout the complete audio range. Because the noise appearing with the signal consists of high frequency components, it is only necessary to perform the pre- and de-emphasis function on the higher audio frequencies.

COMMAND RECEIVER. The circuits of a *Command Receiver,* shown in the functional block diagram, are those of a conventional FM communications receiver with some added refinements. Ordinary superheterodyne principles are used in the second detector. At the limiter stage preceding the detector, the limiter grid current, which is proportional to the strength of the incoming carrier, is used to operate a *carrier-fail relay* in case the carrier should become too weak.

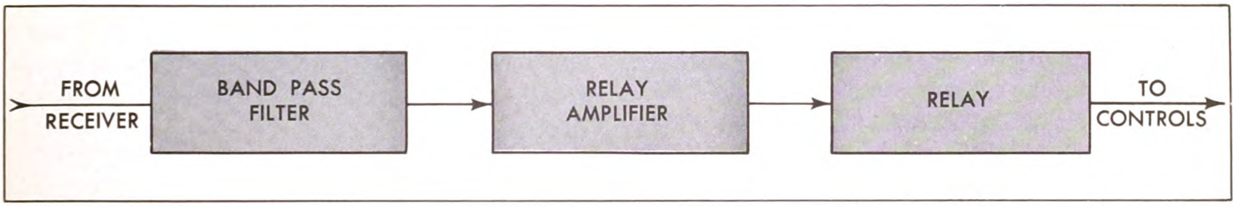
The transmitter carrier is usually left on, even if no modulation is being transmitted, so that the receiver will have a means of determining whether or not it is continuing to receive a signal.

AF Channel Selectors receive the transmitted information after amplification and detection. It is these selector channels and their operation that make the receiver unique in its application. There is one selector channel for every tone that the transmitter may send.

A Single Channel Selector using a band-pass filter is illustrated in the accompanying diagram. A selector tube in the relay amplifier is biased below cutoff and held there until a tone to which the input filter is tuned appears. Upon reception of the proper tone, it is passed by the filter to the grid of the relay amplifier causing a plate current to flow. The plate current flowing in the relay coil energizes the relay which completes the circuit for performing the desired function.



Command Receiver



Single Channel Selector, Gate, or Commutator Segment

SECTION A, CODING AND DECODING, SYSTEMS UTILIZED IN P/I GUIDANCE

An important function in some P/I guidance systems is that of coding (modulating) and decoding (demodulating) information pertaining to flight. Enemy countermeasures or guidance of more than one P/I at a time make coding very necessary.

One of the main disadvantages in the use of a command guidance system is that it limits the number of P/I that can be launched within a given interval of time. This may be very important in a tactical operation. Each P/I must be tracked and controlled separately from the control station. Signals intended for one particular P/I might easily be received by and change the path of some other P/I. However, by using different command signal frequencies, two or more P/I's may be launched and controlled simultaneously from the same control station. But even this technique does not permit an unlimited rate of fire.

A further disadvantage of these systems is their susceptibility to jamming. As previously stated, command guidance systems will generally employ some type of radio or radar link. Countermeasures designed to jam a command guidance system will most likely be encountered near the target area. From the

standpoint of the intruder's defense, such measures require that the signals transmitted to the P/I receiver be capable of overriding the command signals. This, of course, requires a knowledge of the frequency used in the guidance system of the attacking P/I, and the time required to initiate jamming countermeasures. In general, for command systems, a wide range of frequencies must be made available. Allowance must be made for control frequency changes on short notice, and coding systems must be used so that the possibility of jamming the command system will be improbable.

Modulation or coding is a process of impressing intelligence upon a carrier by altering its amplitude, frequency, or phase in accordance with the variations in data to be transmitted. The carrier may be a direct current, an alternating current, or a series of uniform pulses repeated at a uniform rate. You will remember an unmodulated carrier itself conveys no intelligence other than that the transmitter is operating. When some characteristics of the carrier are used to vary as a function of the instantaneous value of the modulating signal, the receiver can detect the variation (demodulate the carrier) and translate it into an intelligible form.

PULSE MODULATION

Increasing demands on military communication services have led to the development of new types of radio and radar systems which have, in turn, necessitated new methods of modulation for the transmission of intelligence in varied and complex forms.

Pulse modulation, with its many variations, has proven most practical in many applications. Fundamentally, pulse modulation differs from other more conventional types in that the intelligence to be transmitted is sampled during brief, periodic intervals, and these samples are used to modulate a subcarrier. The subcarrier is varied in some manner in accordance with the instantaneous value of a modulating signal at the moment of sampling. The subcarrier may be a chain of pulses which are uniform in nature and generated at a fixed rate. An RF carrier is then modulated by this subcarrier pulse chain.

Pulse modulation is valuable for applications which require *multiplexing* or simultaneous transmission of more than one data signal on a common RF carrier. Pulse modulation permits the transmission of many types of data in short periods of time and with a minimum of equipment.

Simultaneous transmission of multiple channels is accomplished by one of two systems. One system involves using a separate subcarrier frequency for each channel, and the other involves transmitting data samples from each channel in a specific time sequence.

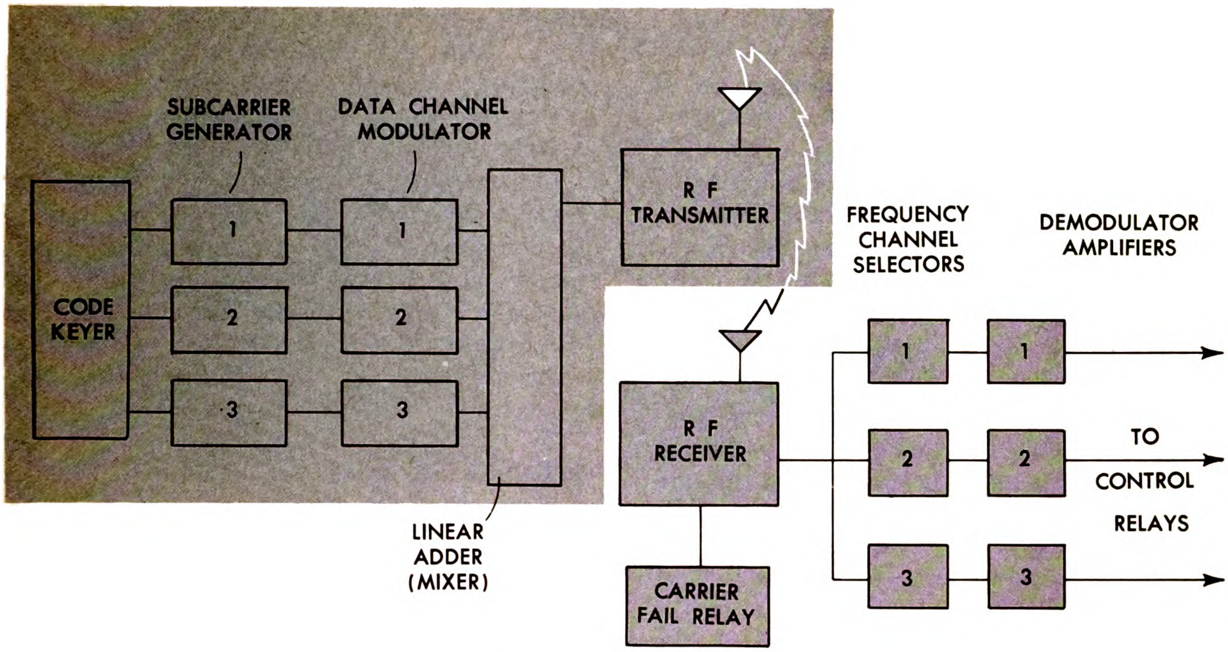
Frequency-Division

In this system, a separate subcarrier frequency is used for each channel. Each data signal modulates the subcarrier assigned to its specific channel and is identified by the frequency of the subcarrier.

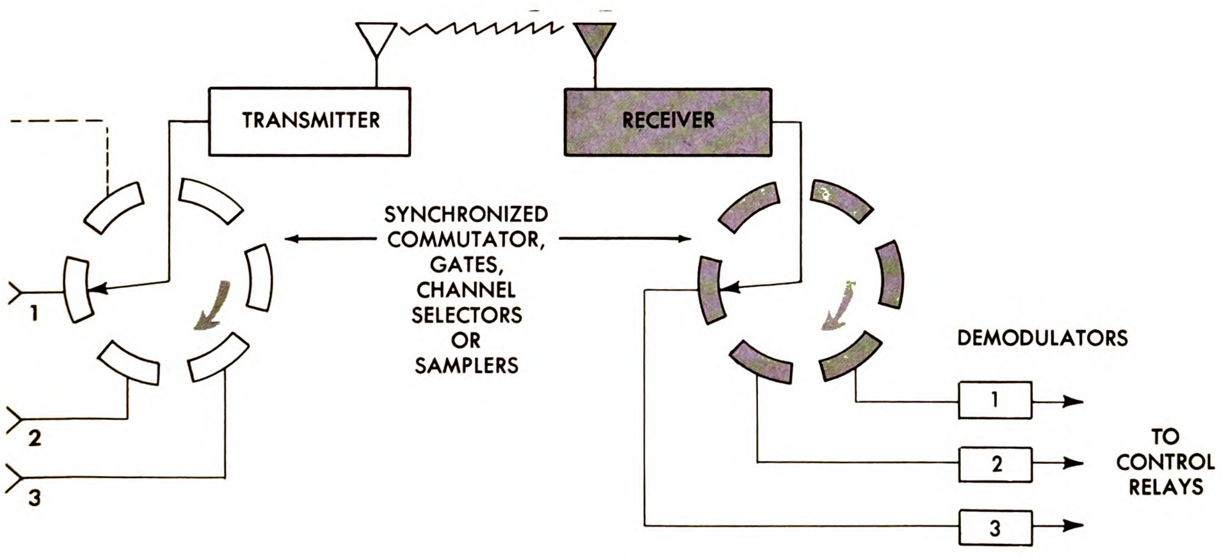
Investigate the *Fundamental Block Diagram of a Frequency-Division System* illustrated. The subcarrier frequencies are modulated by command data in each channel; they are then added linearly into one composite signal train which in turn modulates an RF carrier for transmission. At the receiving end, after the modulated RF carrier has been detected, the modulated subcarrier frequencies are channeled out by the linear frequency channel selectors to their corresponding demodulating amplifiers, where they are demodulated and fed to the proper control circuits.

Time-Division

An alternate system, known as *Time-Division Multiplying or Commutation,* is illustrated in the next functional diagram. Here the instantaneous amplitude of a data



Fundamental Block Diagram of a Frequency-Division System



Time-Division Multiplexing or Commutation

signal is sampled by the transmitter commutator in cyclic serial sequence—from one channel at a time—and a modulated pulse train is generated for each channel. These modulated pulses are then fed to the RF transmitter where they modulate the RF carrier. The pulse train modulated RF carrier is, in turn, transmitted in a corresponding sequence until all channels have been sampled (commutated); the process is then repeated in the same sequence until all desired command data has been transmitted.

Here again, after the pulse train modulated RF carrier has been detected within the receiver, the modulated pulses are fed to the receiver commutator which is synchronized with the transmitter commutator. The selected pulses from the receiver commutator are integrated, interpolated, and in turn programmed to their corresponding relays and control networks according to the coded command guidance data transmitted.

The nature of the data transmitted determines the bandwidth required for transmission and the required frequency of sampling. Since only one instantaneous value of the modulating signal is transmitted through one channel at any given instant in the time-division multiplexing system, there is no cross-talk or interchannel modulation. This interchannel modulation might occur in the freauency-division system because of the nonlinear frequency response of the modulating and demodulating amplifiers.

Pulse-Time Multiplexing

Pulse modulation lends itself readily to multiplexing by time-division because it must employ instantaneous sampling. When pulse modulation is combined with time-division the system is known as *pulse-time multiplexing*. This system possesses many characteristics which are desirable in a communication system for handling complex data. Some of these characteristics are:

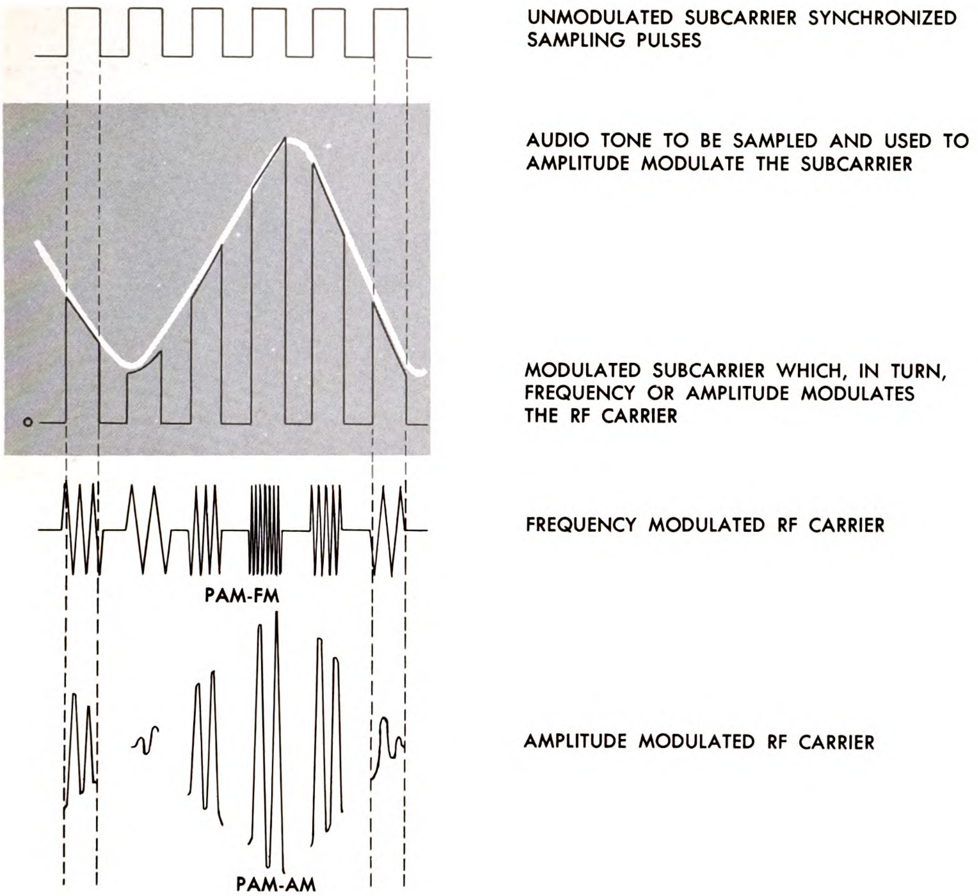
1. High signal-to-noise ratio which is made possible through the use of limiting and clipping circuits.
2. The on-off nature of pulse modulation makes it adaptable to simple repeater or beacon systems for increasing range of transmission.
3. The need for complex filter networks is eliminated.
4. Freedom from interference due to interchannel cross modulation.

A subcarrier pulse chain may have individual characteristics with respect to width, height, duration, repetition rate, formation time, shape, and displacement from normal occurrence. The above characteristics may be utilized singly or in combinations to provide a wide variety of pulse modulation methods. For example, PAM-FM means time-division multiplexing with subcarrier pulses amplitude-modulated in accordance with command data in each channel, and with the subcarrier pulses frequency-modulating the RF carrier. In a PAM-AM system, the subcarrier pulses amplitude-modulate the RF carrier.

Pulse-Amplitude Modulation—PAM

In pulse-amplitude modulation the unmodulated subcarrier consists of a series of regularly recurring rectangular pulses of constant amplitude, width, and PRF. This series of subcarrier pulses (synchronized sampling pulses) is used to sample the audio message (modulating signal) at regular intervals, so as to convert the message into a series of amplitude-modulated pulses. Hence, the amplitude of the subcarrier sampling pulses is varied in accordance with the modulating audio message as illustrated in the *Graph of PAM Waveforms*. Note that the command message to be transmitted is represented by the *white* envelope of the modulated subcarrier sampling pulses.

SINGLE CHANNEL PAM SAMPLING. An elementary method of accomplishing PAM is depicted in the accompanying functional diagram. Here, as indicated, the audio tone is combined with the unmodulated subcarrier sampling pulses and applied to the control grid of a negatively biased cathode follower. V_1 is biased sufficiently below cutoff so that application of the audio tone alone will not drive it into conduction. Hence in the absence of the unmodulated subcarrier pulses, the cathode follower output is clamped at zero. However, when the subcarrier pulses and the corresponding instantaneously sampled components of the audio tone are applied simultaneously to the grid of V_1, the plate current is modulated in



Graph of PAM Waveforms

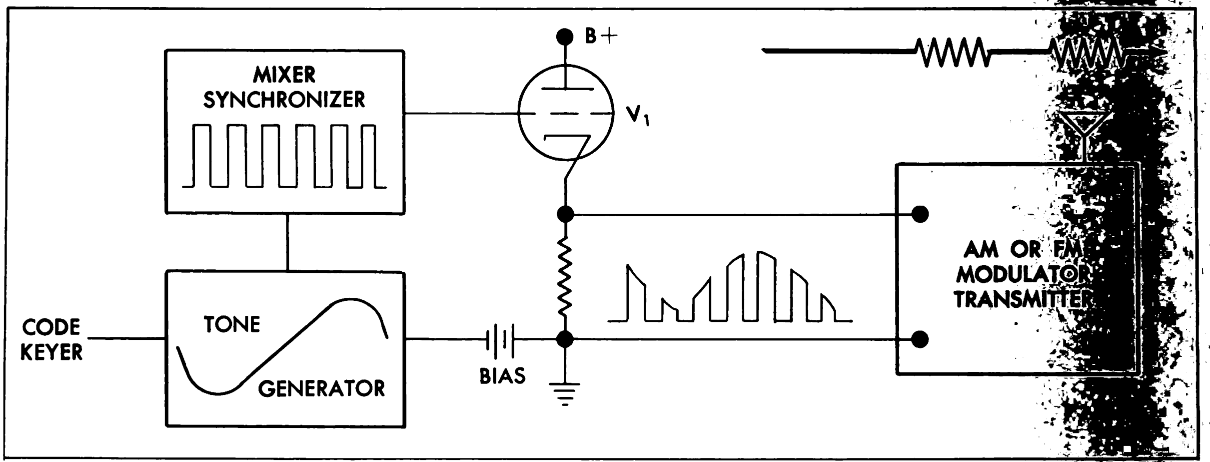
accordance with the instantaneous sum of the two input components. Hence, the magnitude of the cathode follower output varies in direct proportion to the amplitude of the modulating audio tone.

The resulting amplitude-modulated subcarrier pulses are then fed to either an AM or FM radio modulator transmitter—as indicated in the functional diagram.

An actual system may consist of several such PAM sampling channels, where various tone-coded command signals are keyed to amplitude-modulate the subcarrier sampling pulses of their corresponding channels. The modulated subcarriers are multiplexed and transmitted to the command receiver in the form of an amplitude- or frequency-modulated RF carrier.

The amplitude- or frequency-modulated bursts of RF energy are detected in conventional RF command receivers. Here the rectangular subcarrier pulse trains are reproduced and routed to the proper channel equipment by synchronized commutators, gates, channel selectors, or samplers. The PAM subcarriers contained in various channels are then demodulated—converted back into their original audio tone form.

PAM DECODING. In order to extract intelligence from any coded system, it is necessary to develop a synchronizing key or code at the receiving end which will properly match the coded transmission. All types of coded transmission require a decoding circuit that reproduces the original modulating tone and keeps interference and noise at a minimum.



Single Channel PAM Sampling Circuit and RF Transmitter

In the PAM system, as in all amplitude-modulated systems, the amplitude-modulated pulse trains are easily demodulated with a low-pass filter circuit similar to that shown on page 5-12. (See also page 4-37.)

Pulse-Width Modulation

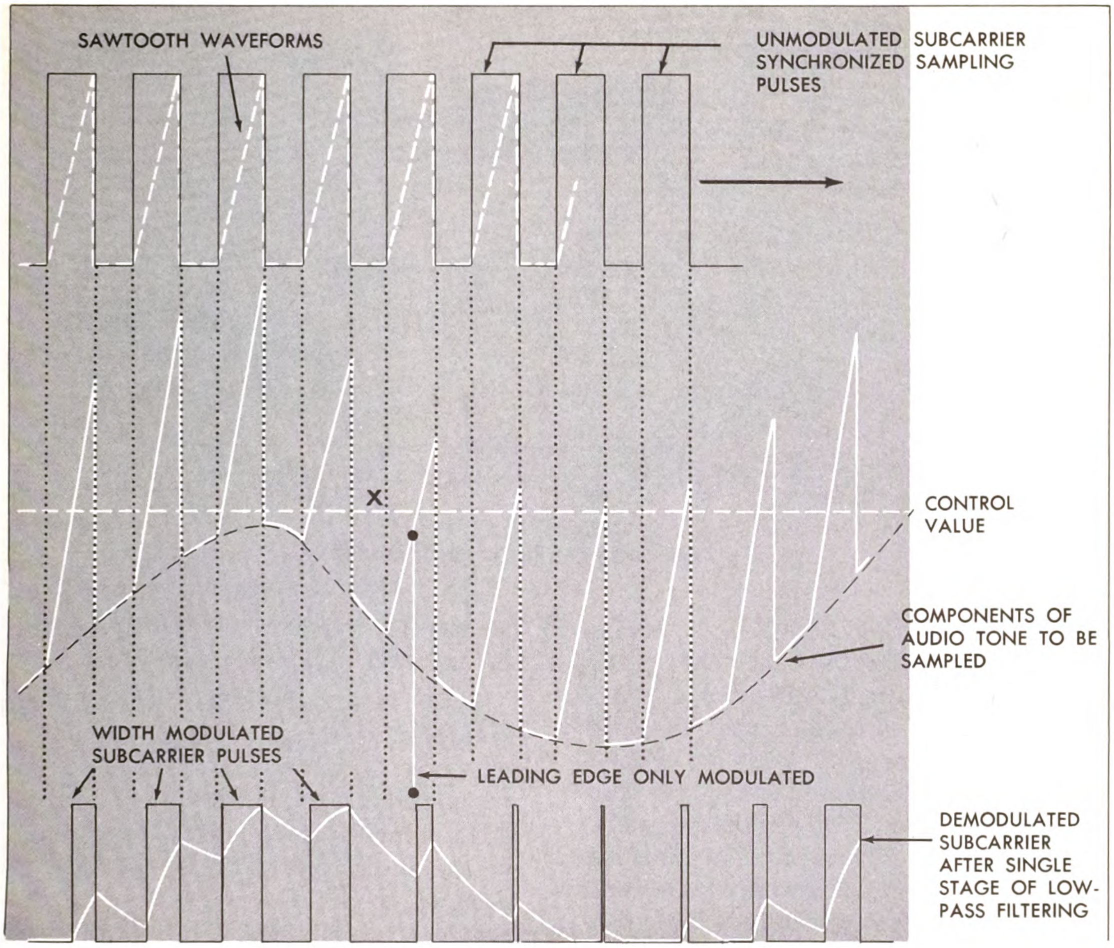
Pulse-width modulation (PWM)—sometimes referred to as pulse-duration modulation (PDM) — is actually a type of pulse-time modulation (PTM). However, instead of varying the relative position of the subcarrier sampling pulses, individual instantaneous samples of the modulating audio tone cause their corresponding subcarrier pulses to vary in width or time duration.

A SINGLE CHANNEL PWM SYSTEM. The primary waveforms of an elementary single channel pulse-width modulated system are illustrated. The uppermost waveform represents an unmodulated frame of subcarrier sampling pulses. These rectangular pulses are converted into sawtooth waves by synchronized start-stop linear integrators (see Integrator Components of Computers as treated in the previous chapter). The sawtooth waveforms are then added to the sampled intervals of the modulating audio tone as shown in the second group of waveforms. The resulting composite waveform is next applied to the input of a slicing circuit which has the ability of producing a pulse of constant amplitude when its combined input rises above a predetermined control value.

The black rectangular pulses in the lower group of waveforms represent the output of the slicer, the pulse-width modulated subcarriers. Note that the output of the slicer increases from zero to its constant value at the instant the sawtooth component of input exceeds the control value, and that the output suddenly drops to zero whenever the trailing edge of the sawtooth becomes lower in amplitude than the control value.

Now re-examine the group of PWM waveforms and note that the trailing edges of the modulated subcarrier pulses are fixed (unmodulated), and the leading edges vary in accordance with the modulating audio tone. In other words the leading edges alone are modulated. Hence, the width of each modulated subcarrier sampling pulse is proportional to the instantaneous magnitude of the audio tone at the instant of occurrence of the variable edge of the pulse—as at point x in the illustration.

The subcarrier pulses of varying width are used to pulse-modulate the RF carrier. A coded tone is thus transmitted to the command receiver in the form of width-modulated bursts of RF energy. The bursts of



Primary Waveforms of an Elementary Single Channel PWM System

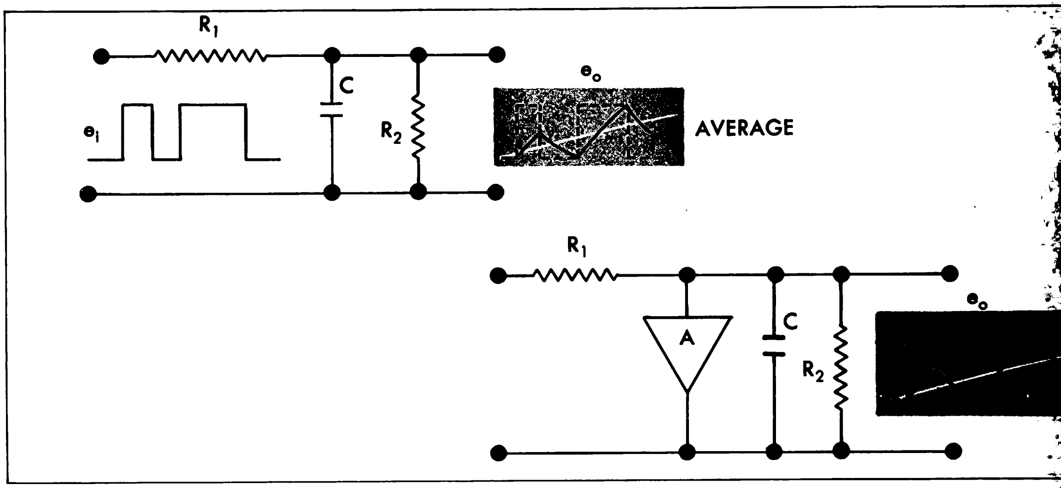
RF are detected and demodulated by command receivers, where the rectangular subcarrier pulses of varying widths are reproduced.

DEMODULATION OF A PWM SUBCARRIER. Recovery of the original audio tone from a pulse-width modulated train of rectangular pulses may be accomplished by passing the pulse chain sequentially through a low-pass filter integrator.

A Single L-Section Low-Pass Filter and its analog form are depicted in the accompanying diagram. When the subcarrier pulses are impressed upon the input to the filter, the integrating capacitor charges through R_1. During intervals between pulses of varying width, the capacitor discharges through R_2. Hence, the magnitude of the resulting audio tone depends upon the time constant of the RC circuit as well as the magnitude of the subcarrier pulses. The pitch of the audio tone depends upon the rate and accumulation of charge which are proportional to the product of RC, width (charging interval), and resting time (discharging interval) of the duration-modulated pulses.

Pulse-Time Modulation

Pulse-time modulation is often referred to as *pulse-position* or *pulse-displacement* modulation. This system of coding is widely used in microwave time-division multiplexing because it permits interlacing several subcarrier pulse chains without confusion.



Single L-Section Low-pass Filters

Pulse-time modulation (PTM) is accomplished by varying the time between pulses or by varying the displacement of a signal pulse with respect to a reference (marker) pulse. A marker pulse is supplied from a separate marker generator, such as a free-running multivibrator. The marker generator modulates an RF carrier at uniform intervals with pulses of constant amplitude and width, or with pairs of pulses which are readily distinguished from those produced by the signal.

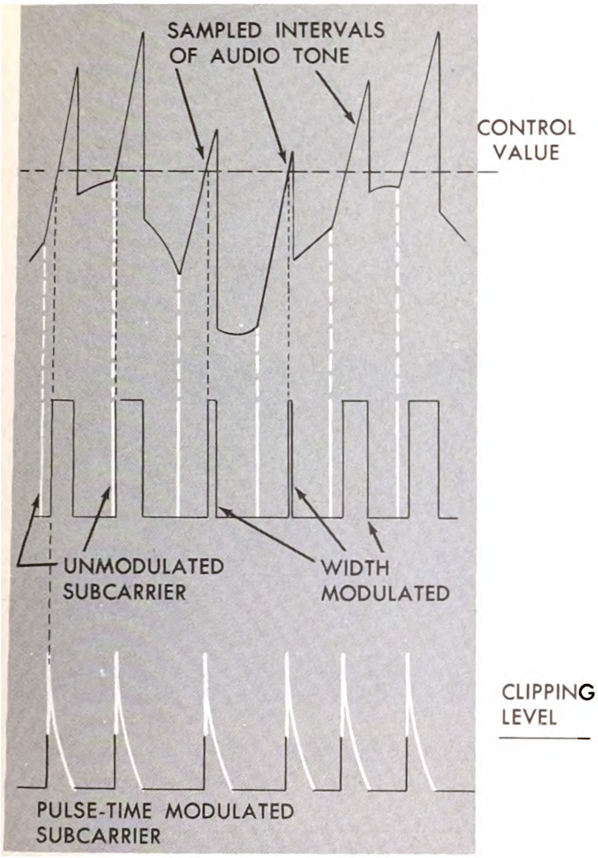
MARKER PULSES. Marker pulses coded with a predetermined repetition period are used for synchronization of both transmitting and receiving equipment. For example, a commutator or channel selector may consist of a synchronized series of gates which separate individual pulses within a pulse chain or frame and route them to predetermined channel equipment. That is, by timing the open and closed intervals of the channel selector gates, only a particular channel is open when predetermined pulses that should pass are being received.

Synchronizing equipment, at the transmitting end, controls the generation of marker pulses as well as the time of sampling to produce the predetermined modulated subcarrier pulses. Paired pulses may be used for markers so that they [illegible] the message carrying pulses.

PTM MODULATION METHODS. [illegible] several methods by which an audio [illegible] nal can be applied to produce [illegible] modulation. One method is to [illegible] from a pulse-width modulated cha[illegible] is, the audio tone to be transmitted [illegible] sampled to produce a PWM pulse [illegible] previously explained. The width-modulated pulses are then fed to a short time constant differentiating circuit, the output of which is next fed to a limiter and clipping circuit. Hence, there results a series of pulses whose amplitude and width are constant, but the position of the pulses varies in accordance with the magnitude and frequency of the audio tone, as depicted in the waveforms in the accompanying diagrams.

Blocking Oscillator Method of [illegible] *PTM* employs a driven blocking oscillator which conducts when the positive-going [illegible] tion of an audio signal is applied to [illegible] trol grid or when the negative-going [illegible] is applied to its cathode.

Hence, positive and negative [illegible] an audio signal may be applied [illegible] to the grid and cathode of a single [illegible] oscillator through limiting or [illegible]



Deriving PTM Through Differentiation, Limiting and Clipping a PWM Pulse Chain

cuits. On the other hand, two properly biased blocking oscillators may be used, one of which is biased to conduct only on the positive portion of an audio signal, while the other is biased to conduct only on the negative portion. In either case, the end result is the same. The blocking oscillator conducts and produces a modulating subcarrier pulse when the amplitude of the audio signal attains a predetermined positive and/or negative value. Therefore, the frequency of the audio signal determines the frequency at which the modulating subcarrier pulses are generated.

When the generated modulating pulses are superimposed upon or used to trigger a chain of carrier pulses which is being modulated at uniform intervals by a marker pulse (or pair of pulses), the position of the modulating subcarrier pulses with respect to the modulating marker pulses varies in accordance with the frequency variations of an audio signal.

In this manner, the intelligence conveyed is represented in terms of the relative time or position between the pulse-time modulated subcarrier pulses and the marker pulses. *A Pictorial Representation of a Pulse-Time Modulated Subcarrier* is illustrated. Note that the audio-modulating signal is represented by four sinusoidal waveforms, each of a different frequency. Look, for example, at the second sine wave. The frequency of this modulating signal is represented as being equal to one-half the pulse recurrence frequency (PRF) of the *paired marker pulses.* Hence, each half of the modulating cycle produces a pulse which is exactly midway between the adjacent paired marker pulses. These two time-modulated subcarrier pulses are separated by the same time displacement as the marker pulses.

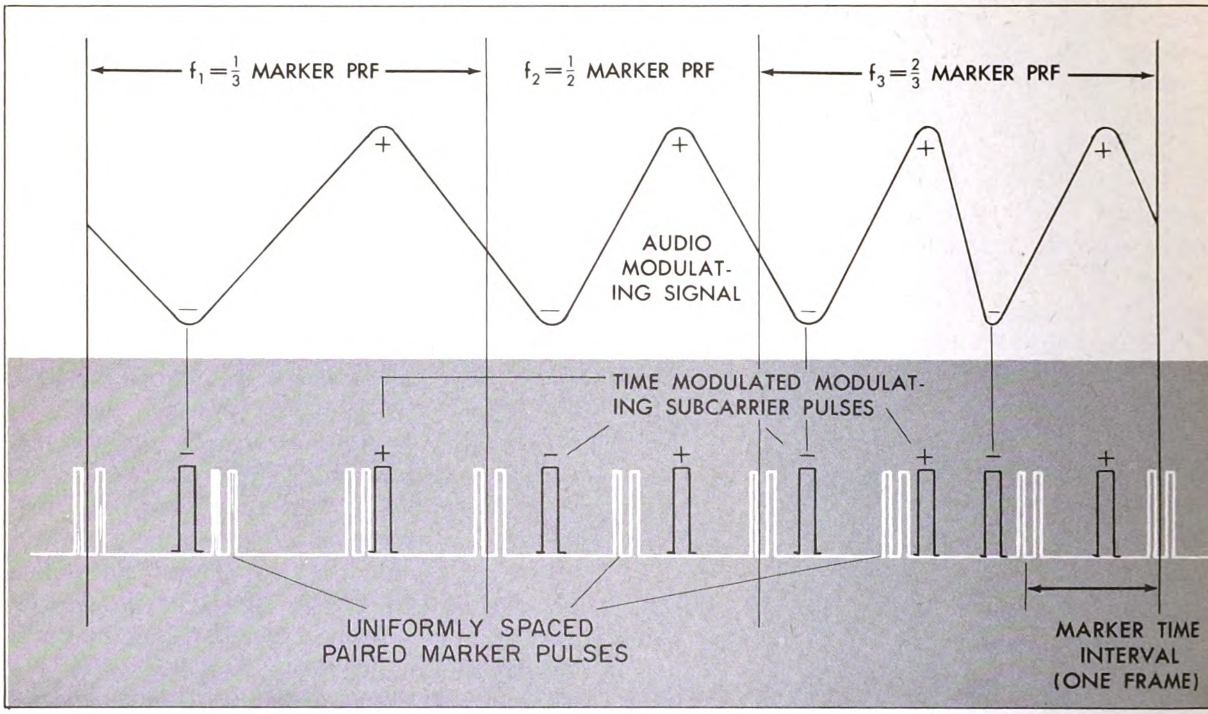
The first sine wave represents a modulating signal frequency equal to one-third that of the paired marker pulses. Hence, the time or distance between the modulated subcarrier pulses is correspondingly greater. No pulse appears between one pair of marker pulses.

The last group of sine waves represents a modulating signal of higher frequency—two-thirds the marker PRF as indicated. The distance between the time-modulated subcarrier pulses is now correspondingly less, with *four* pulses appearing within *three* marker time intervals.

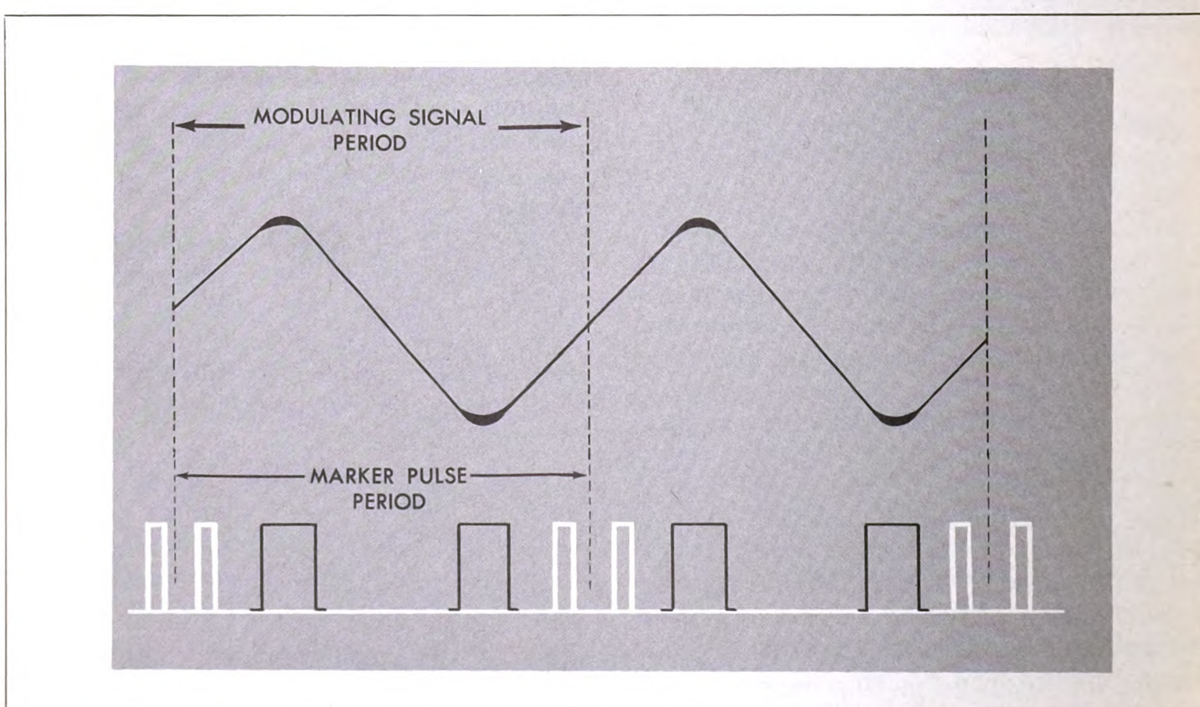
If the audio-modulating signal were the same frequency as the marker PRF, two pulses, one for each half of the modulating cycle, would appear between each successive pair of marker pulses as illustrated.

Pulse-Time Decoding

In the pulse-time modulation system just discussed, three pulses are transmitted by



A Pictorial Representation of a Pulse-Time Modulated Subcarrier



PRF of Paired Marker Pulses Equal to Modulating Signal Frequency

radar. The first two are coded for beacon response. The third pulse is shifted in time according to the intelligence being transmitted. At the receiver, the first two pulses (perhaps a pair of marker pulses) are decoded and used as a reference and for beacon triggering and synchronization. A contemporary command guidance system employs a radar that supplies the three necessary pulses, with the third pulse variable in time to provide the information for proportional control. A *Basic Pulse-Time Discriminator* utilizing a coincidence circuit and delay time, along with its accompanying waveforms, is shown in the next illustration.

The cathode follower stage is necessary to provide a good impedance match between the input circuit and the coincidence and delay line circuitry. In this example, the leading edges of the input pulses to be detected are spaced 3.5 microseconds apart, and the pulses themselves are 0.5 microseconds wide. This signal is fed to the control grid of the first tube in the coincidence circuit and to the input terminals of the delay line. The delay line is tapped in such a manner that the signal is delayed by an amount exactly equal to the spacing of the two pulses; the signal is then fed to the grid of the V_2, the second tube in the coincidence circuit.

As you well know, any pulse applied to either grid will affect the current flow through the common plate load resistor R_L. For example, when the first pulse (A) is applied to the grid of V_1, R_L passes more current causing the plate voltage to drop—producing a negative pulse at the output, as indicated. The second pulse (B), when applied to the grid of V_1, also causes a negative pulse to appear at the output. However, at exactly the time that pulse (B) is applied to the grid of V_1, a delayed component of the first pulse (A^1) appears at the grid of V_2 augmenting the drop in plate voltage.

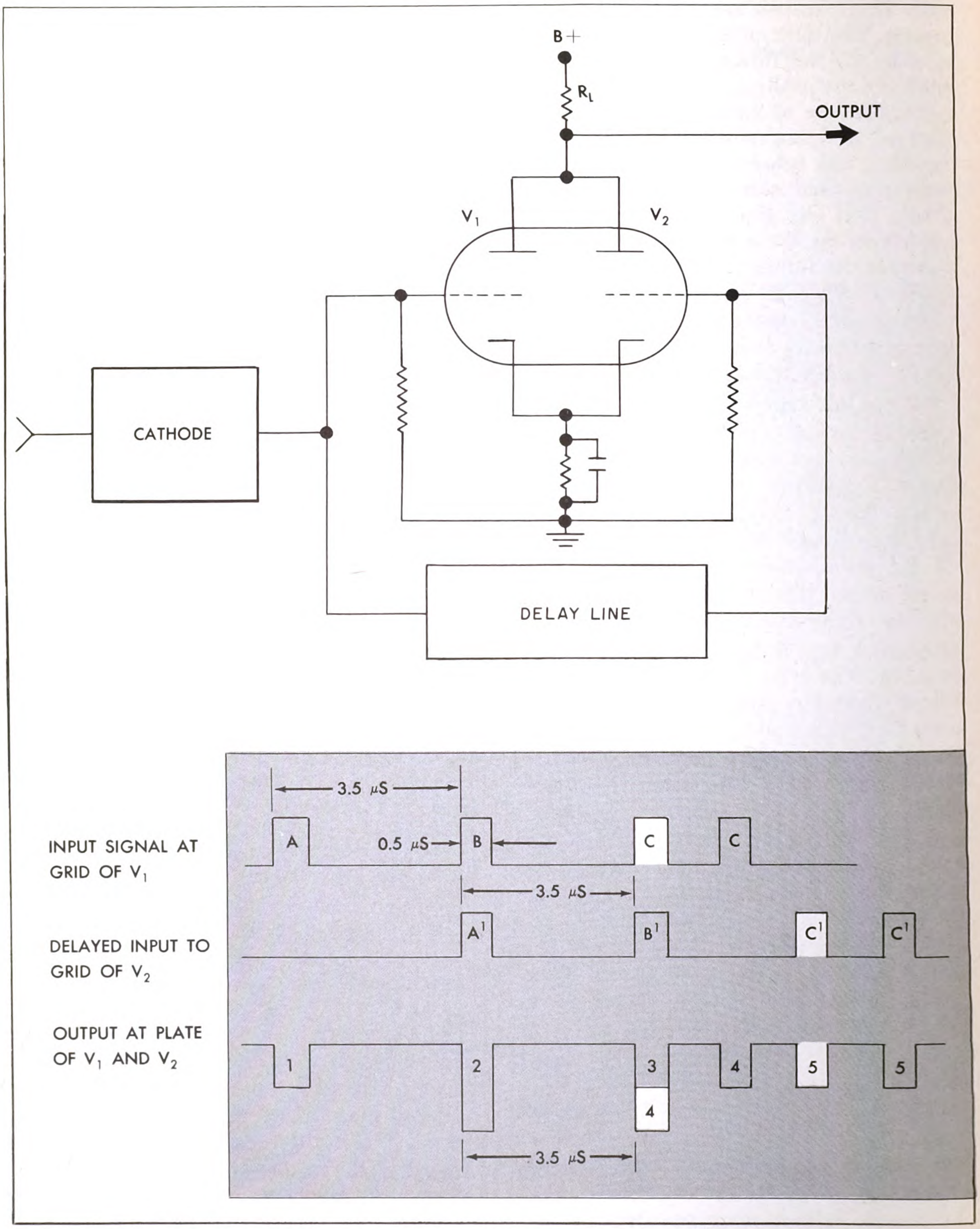
When the delayed component of the second pulse (B^1) appears at the grid of V_2, it causes a negative output pulse of considerably lower amplitude, because the third input pulse is not in coincidence with B^1. If the third input pulse (C) has appeared at the grid of V_2 in coincidence with B^1 at the grid of V_1, the magnitude of the third output pulse would have been equal to the second.

The resulting output pulses may be used to gate a number of different circuits to indicate that pulses of a particular spacing have been received at the input. It must be remembered, however, that any circuit to be triggered by the output of the coincidence circuit must be biased so that only an output pulse, resulting from the two input pulses appearing simultaneously on the grids of V_1 and V_2, would have sufficient amplitude to trigger that circuit. Therefore, a pulse applied to only one grid (V_1 or V_2) will not yield an output with magnitude enough to trigger the next circuit. The pulses which are to be applied to the input of the discriminator must, of course, be uniform in amplitude. This uniformity of amplitude may be accomplished with a limiter stage.

One-Radar Command Beacon System

The use of a one-radar system for both tracking and transmitting command signals forms a rather unique combination. Such a combination may be utilized to extend the tracking range for relatively small P/I which might ordinarily offer a poor radar return. Consider, for example, installing a radar beacon in a small P/I.

When properly triggered, the beacon transmitter will transmit a much stronger pulse back to the ground based radar transmitter than the echo that would normally be returned. The beacon in the P/I is a small receiver and transmitter operating within the ground based radar tracking band. The receiver section of the beacon has circuitry that will accept only those signals from the tracking radar which have a definite pulse separation. If the proper coded pulse combination is received by the beacon



A Basic Pulse-Time Discriminator

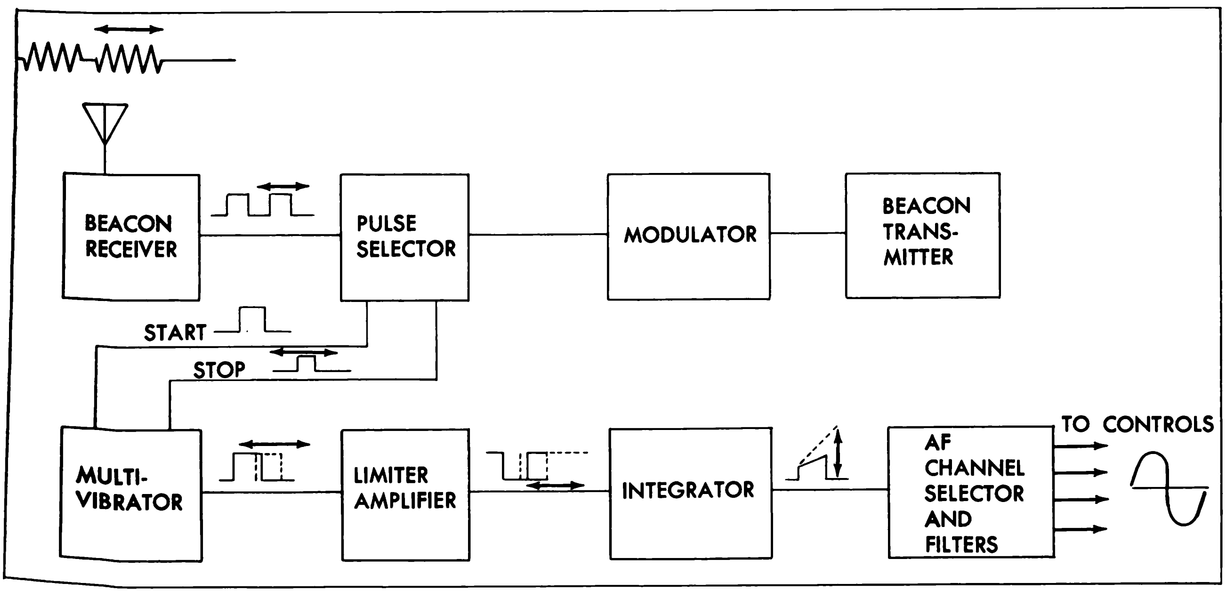
receiver in the P/I, the beacon transmitter will send out a pulse. The beacon transmitter does not operate on exactly the same frequency as the tracking radar, but on a slightly different frequency. The tracking radar receiver is also tuned to this frequency. The reason for this is to permit the tracking radar receiver to distinguish the signal transmitted by the radar beacon from the reflected return of the P/I tracking radar.

In addition to its functions relative to tracking, the radar beacon receiver in the P/I may be designed to accept command signals from the P/I tracking radar. To accomplish guidance, an additional set of coded pulses is added to the signal transmitted by the ground based tracking radar. The arrangement of these pulses determines the information contained in the command signal. The radar beacon receiver in the P/I accepts these coded command signals and routes them through a network which decodes them for the guidance information they contain. This information is impressed on the signal from the ground radar as pulse-position modulation (PPM) and gives the P/I its guidance commands.

Pulse-position modulation requires *two pulses per sampling cycle to give one intelligence channel.* As previously stated, the output of the ground radar contains both the P/I beacon interrogation pulses and the command signal pulses. The position of the modulated pulse is varied back and forth about its rest position at an audio rate. Consequently, the number of intelligence channels is limited by the audio-modulation which can be accomplished by the sampling frequency. The sampling rate depends on the operating range of the radar set.

COMBINATION RADAR BEACON AND COMMAND RECEIVER. Let us now examine the diagram and see how it is possible to use the beacon as a command receiver. The beacon receiver converts the pulsed RF signals into a set of video pulses which are separated by the pulse selector. The first pulse is used to start, and the second pulse is used to stop the multivibrator, thus producing a square wave of varying width.

The varying width of the multivibrator output follows the excursions of the modulated pulse. Subsequently, an integrator forms a sawtooth waveform whose amplitude



Combination Radar Beacon and Command Receiver

varies with the width of the square wave. In this manner, the pulse position or time difference in the received signal is converted to amplitude variations of a sawtooth wave whose frequency is the same as the radar PRF. The sawtooth waveforms are fed into a frequency selective amplifier (filter) where all components except the fundamental modulating frequency—which is evidenced by the amplitude variations—are removed. The resulting sinusoidal output pulse then excites a relay amplifier and initiates the desired control function.

Beam-Rider Command Guidance

The theory of operation of beam-rider guidance is based upon the principles of automatic tracking radar systems and has its primary use in air-to-air (AAM) and surface-to-air (SAM) P/I. Let's consider a P/I designed to be guided and controlled by a pulse-modulated radar system. Detection of an intruding target in such a system may be accomplished by directing a beam of high-frequency radio energy pulses in a predetermined position over the area to be searched.

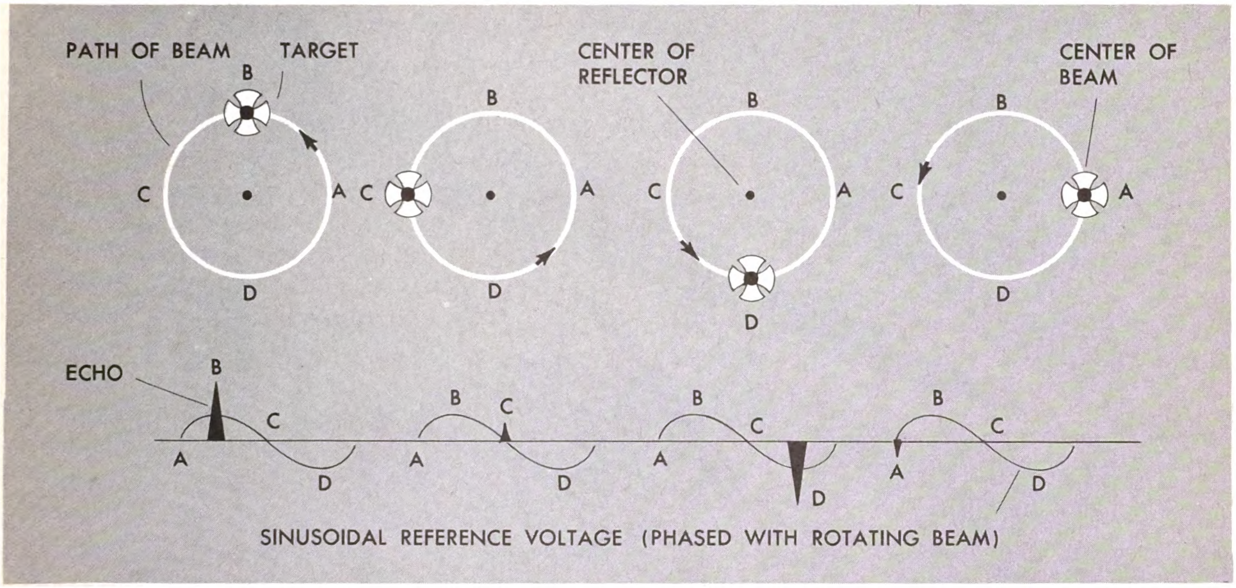
When the radar beam strikes an object, energy is reflected back to the transmitter. A small portion of this reflected energy is detected by a receiver which is a part of the radar system. Due to the design of the transmitting antenna, the beam is narrow, so that when reflections (echoes) are received, the direction of the target can be determined. By moving the antenna until the echoes are of maximum strength, the target will be found upon the axis of the antenna, or in other words, on the center of the beam. Since the beam is pulsed, the elapsed time between the transmission of a pulse and the return of its echo provides a means for determining the range of a target. The automatic tracking radar generally employs a dipole antenna with a parabolic reflector. The same antenna can be used for transmitting and receiving by using what we call a duplexer or TR box.

CONICAL SCANNING. You will find that in automatic tracking radar systems, the trans- mitted beam is not fixed with respect to the axis of the antenna reflector, but is [illegible] able so that the beam will literally [illegible] cone in space. This is accomplished by [illegible] tion at either the radiating dipole [illegible] antenna reflector. By rotation we mean [illegible] spinning of the dipole which is offset [illegible] from the axis of the reflector to produce [illegible] conical beam pattern. However, the [illegible] may be stationary and the reflector [illegible] and rotated to produce a similar [illegible] beam pattern. These procedures of ob[illegible] a conical beam pattern are called [illegible]

Nutation of radar antennas [illegible] efficient conical scanning. The reflected [illegible] nal is always received in greater strength because it has the same polarization as the transmitted signal. However, in the rotating antenna system, the polarization of the echo signal will always lag the antenna position slightly and the received signal will be weaker than it would be if its polarization exactly matched that of the antenna.

A target located on the axis of the reflector always produces an echo of constant amplitude for all positions of the beam, because the beam describes a uniform path around the reflector axis. If the target moves away from the reflector axis as the beam rotates, the echo signal varies in an approximately sinusoidal manner. Thus, the target direction may be determined by comparing the phase of the echo signal with a reference voltage.

From *The Target Position Identification Diagram* illustrated, we see that this error voltage, obtained by comparing the echo signal with the reference voltage, is used to control the antenna positioning system of the tracking radar. The antenna is moved continuously in order to cancel the error voltage and to keep the echo voltage in phase with the reference voltage. The development of position error voltage by means of conical scanning forms the basis of an automatic tracking radar system.



Target Position Identification Diagram

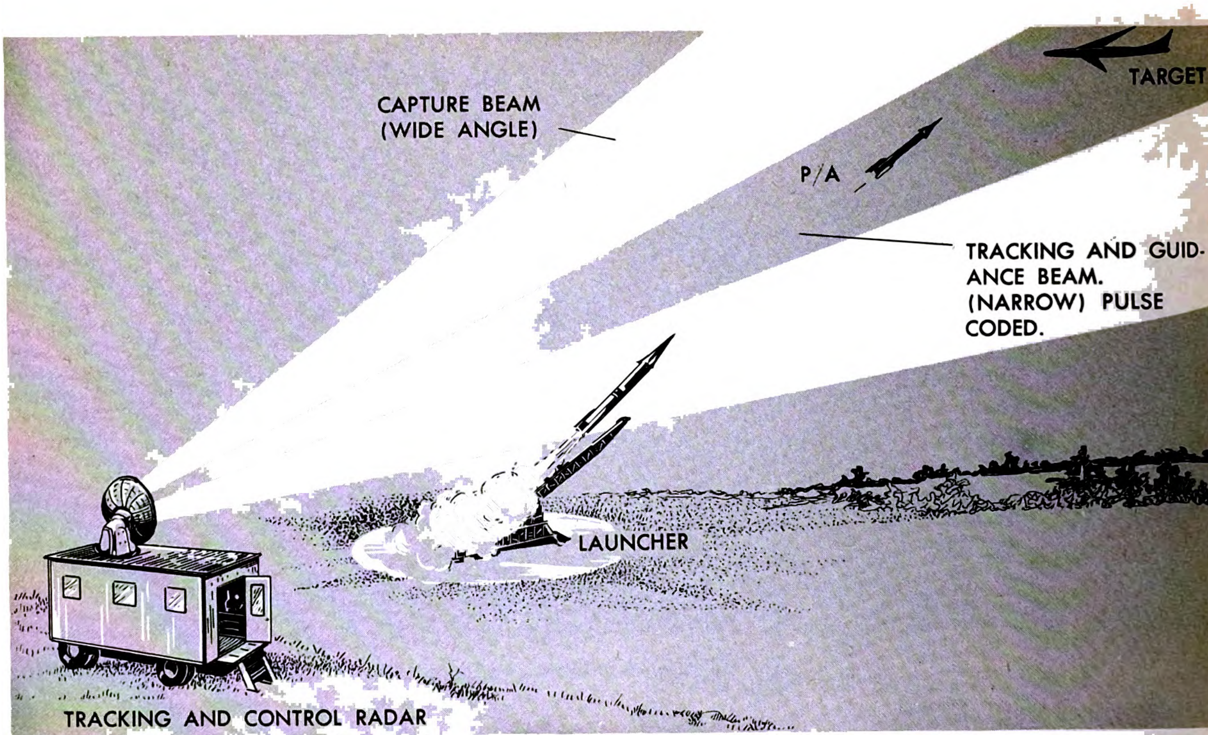
For small deviations of the target, the amplitude variations of the error signal produced by combining the echo voltage and the reference voltage will indicate the amount of displacement of the target or its distance from the reflector axis. The reflector axis or the axis of the conical trace of the radar beam represents a straight line path to the target. Our concern is a method for controlling the P/I so that it will fly along this straight line path automatically, or in other words *ride the beam* to the target. To accomplish this, it is necessary for the P/I to sense when it is on the beam axis, or if displaced, to sense the direction and amount of displacement and readjust its course accordingly.

The P/I has no internal phase reference, as does the radar tracking system. Therefore, the beam must provide a modulated signal to supply the required guidance information to the P/I. The beam may be modulated in such a manner that, in addition to its tracking and range-finding functions, it can supply the P/I with sensing information in a cone-radar system, or it can supply position data signals to the command control radar in a two-radar system.

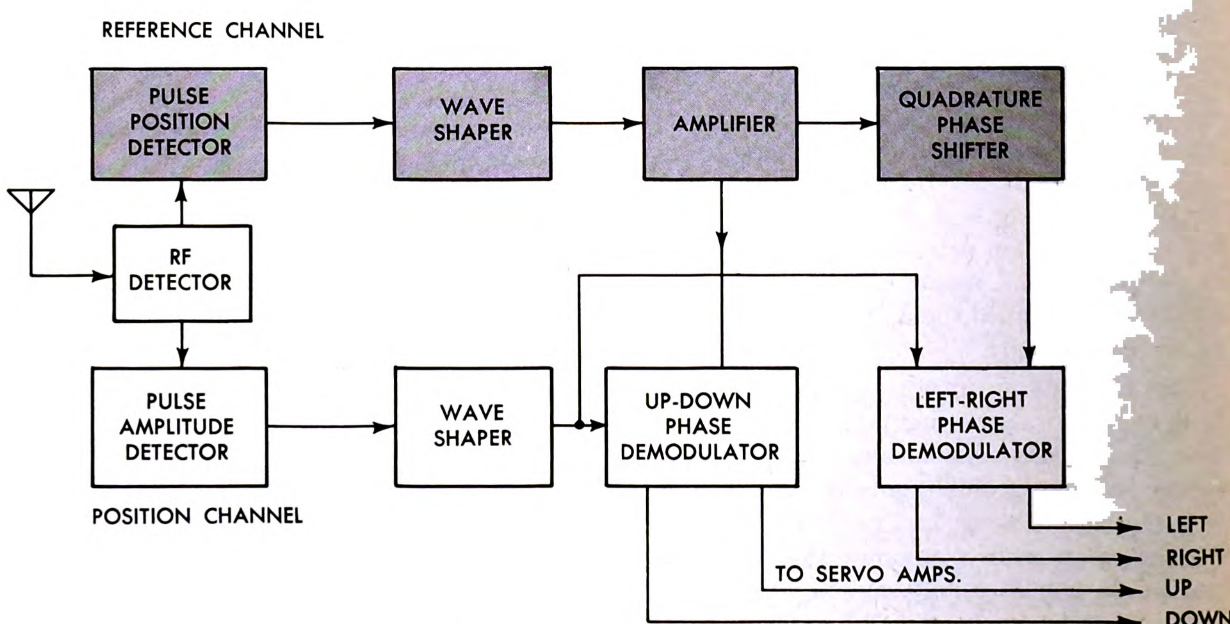
ONE-RADAR SYSTEM. Investigate the *One-Radar Beam-Rider System* illustrated. As indicated in the illustration the rotating beam has signals superimposed upon it (modulation) whereby it can determine the position of the P/I with respect to the axis of the beam. These signals may consist of coded pulses which are transmitted while the antenna is rotating within a specific quadrant.

The coding may be in the form of pulse-time modulation applied to the tracking pulses at specific intervals of time. Then, when they are detected in the pulse position detector of the *P/I Command Guidance Receiver* (illustrated), they will supply a sinusoidal error voltage to the position channel.

This error voltage will have the same frequency as the scanning frequency, which is generally some value between 20 to 120 cycles per second. The reference voltage is supplied from a generator driven by the scanning drive motor and is therefore the same frequency as the scanning beam. This



One-Radar Beam-Rider System



P/1 Command Guidance Receiver

voltage is compared in the position channel phase demodulator with the sinusoidal wave resulting from the amplitude modulation of the received pulses and then phase detected to give the direction of error. The error signal is applied to the P/I control circuits to provide correction of the flight, along either or both the pitch axis and the yaw axis. Correction will continue until the error signal voltage is in phase with the reference voltage which indicates that the P/I is again on the axis of the beam.

In a tracking system using quadrant identification signals, the pulses of the tracking radar are not used to develop position error signals for guidance of the P/I, only the modulation of the pulses is used for this purpose. This system is relatively immune to jamming, as the coding of the pulses by means of the time spacing between them is the only information required. When the pulse-time coding is correct, the pulses will produce conduction in a coincidence circuit in the P/I receiver. This will allow the position signal to pass through to the phase detector. If the position signal is in phase with the reference signal, no error signal will be developed, but if the signals are not in phase, an error will result.

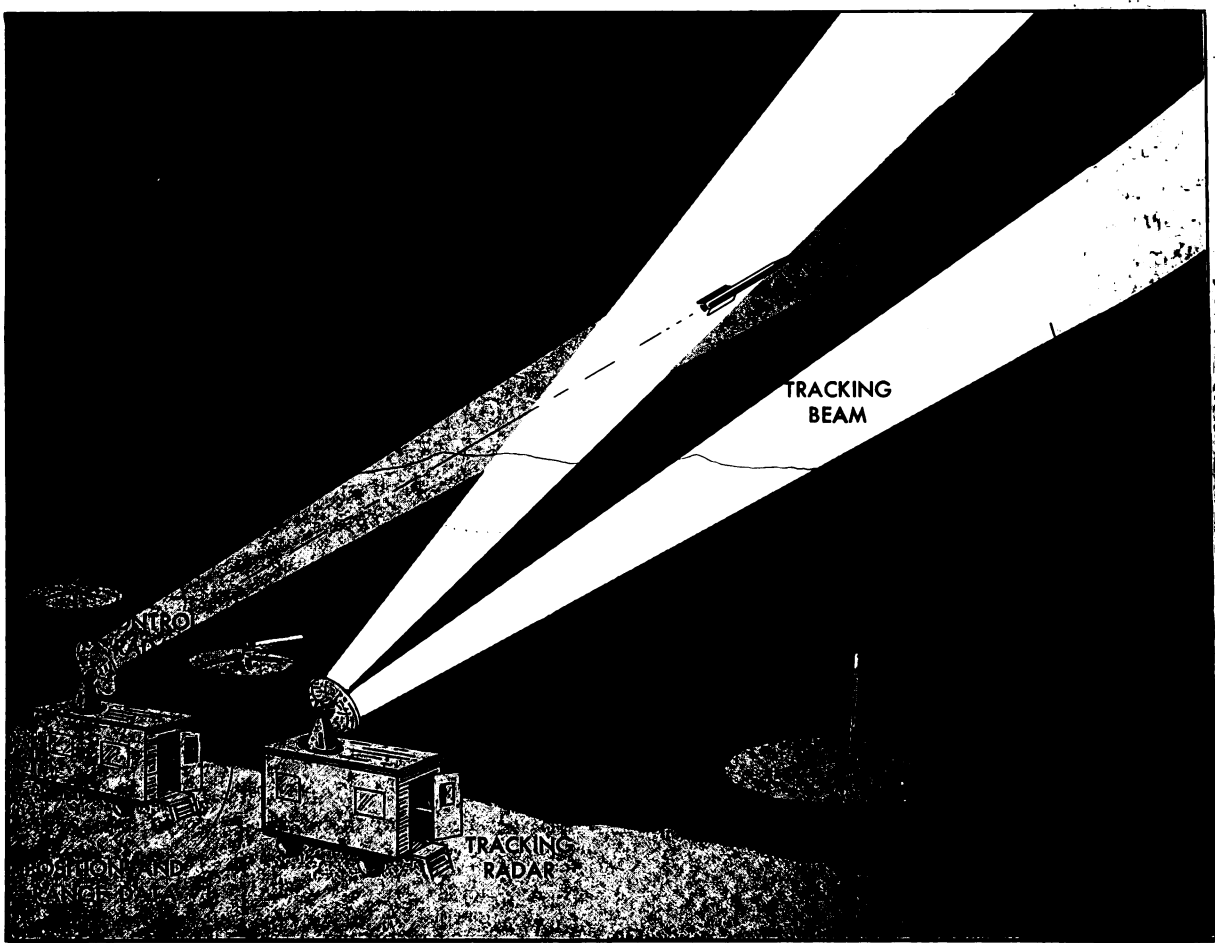
We can consider these phase-compared or quadrant identified signals as commands for up, down, right, and left. As long as the P/I is on the axis of rotation, all signals balance and cancel each other to produce a *no error* output. However, when the P/I is displaced from the beam axis, one signal overbalances the other and the control circuits obey the stronger signal in order to turn the P/I back toward the beam axis. When the one-radar system is used in a beam rider AAM, the launcher aircraft, the AAM, and the target must be kept in line at all times. Even though this condition is difficult to maintain, it is possible with the one-radar system to launch several AAM's in rapid succession and control more than one simultaneously within the area covered by the tracking beam. So you see, the lesser degree of accuracy can be compensated for by directing a greater number of AAM's toward the target.

In the one-radar system wherein the tracking beam serves as the positioning signal carrier and is being rotated constantly, the P/I actually tries to follow the circular path of the beam rather than its axis of rotation. Thus, the P/I is subjected continually to up-down and right-left accelerations which could produce serious aerodynamic stresses and result in an erratic pattern of flight. A system which overcomes many of the undesirable effects and further increases immunity to radar jamming is a two-radar guidance system.

TWO-RADAR SYSTEM. This system, shown in the accompanying illustration, includes a target tracking radar and a P/I control radar. The output of the tracking radar is fed into a *director computer* which points the director or control radar beam at the predicted position of the intruding aircraft.

The P/I may be launched directly into the main beam of the control radar, or it may be launched into the wide-angle *capture beam* and then be directed into the path of the main beam by which it is guided to its target. The pointing of the control beam is continuously corrected, but the movements are very small compared to those of the tracking beam and thereby impose smaller aerodynamic stresses upon the P/I.

The Tracking Radar points the control radar as it tracks the target, and the control radar tracks the P/I and supplies the guidance signals required to maintain its trajectory to the collision point. This two-radar system can handle only one P/I at a time, so the rate of fire is low. On the other hand, it is possible to supply several control radars with pointing data from a single tracking radar and by means of coded control pulses (to eliminate cross-interference) guide several P/I's toward a *target* from separate launching sites. This multi-radar system increases the probability of interception of the target. However, it also increases the requirements for frequency control and dis-



Two-Radar Beam-Rider Guidance System

crimination against interference between the various channels of the system. It also means that component accuracy and reliability be maintained within very close tolerances.

Guidance Position and Reference Signals are received by the antenna of airborne P/I as coded pulses. These pulses are then sent to a guidance receiver wherein they are detected and amplified. After sufficient amplification, the signals are decoded by suitable coincidence circuits. Next, we find they are compared in a phase comparator circuit, filtered, and applied to the P/I control system as up-down or right-left error signals in the form of voltages. By their phase displacement these voltages indicate the *direction* of position error, and by their amplitude indicate *amount* of position error. They are proportional to the deviation of the P/I from the axis of rotation of the tracking beam.

We can distinguish the beam-rider P/I by the location of its guidance antenna, which is designed and mounted in such a manner that it can pick up signals only from sources aft of the airframe. In this way it would not be possible for the P/I to accidentally pick up the echo signals from the target tracking radar. The echo signals could produce a false timing reference, phase cancellation, or noise interference which might affect the control system. Quadrant position identification also could be confused due to the dispersion of the echoes.

SECTION B, CONTROLLERS AND ACTUATORS

In any attitude control system the desired corrective action of control devices is accomplished by a controller-actuator combination. An actuator is the device connected to a control mechanism which supplies the energy necessary to produce the desired corrective displacement.

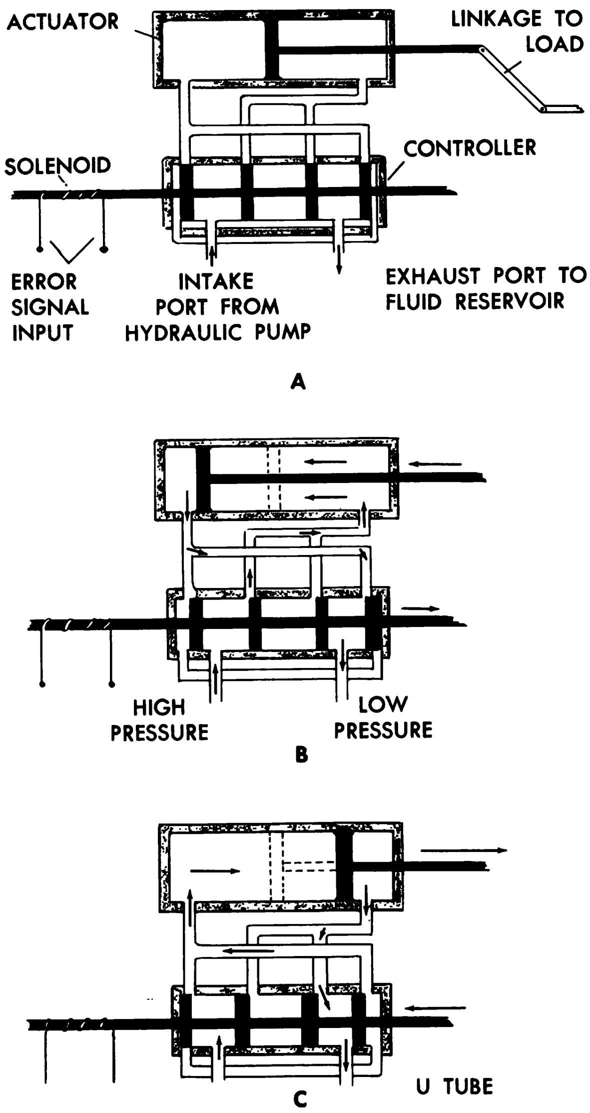
A controller is designed to regulate the *magnitude* and *direction* of actuator movement in response to an existing attitude error signal. For example, an electronic error signal of certain polarity and amount, after transmission by the controller, will produce an actuator displacement of a certain magnitude and in some definite direction. If either the polarity or magnitude of the error signal changes, the direction and magnitude of actuator displacement will also change. Thus, the various controllers and actuators of attitude control systems receive, transfer, and transform the energy originating in the sensing element into mechanical motion of the control devices.

Controller and actuator units are generally classified on the basis of the method, or methods, of energy transfer they employ. The units most frequently used in control systems are either hydraulic, pneumatic, or electric. Combinations of these have also been effectively used.

HYDRAULIC CONTROLLER AND ACTUATOR UNITS

Depending upon the work to be performed, the component parts of a hydraulic actuating system will vary in arrangement and physical dimensions. Generally, a hydraulic controller consists of a valve which is automatically positioned by a solenoid or electric motor in response to an electric error signal, as depicted in the accompanying illustration. As you see, this controller unit is also equipped with a high-pressure intake port connected to a hydraulic pump or accumulator, and an exhaust port connected to a hydraulic fluid reservoir. In addition, the two ports on the output side of the controller valve lead to opposite ends of the actuator cylinder. Each of these ports alternates as delivery and return orifices between the controller and the actuator, depending upon controller valve position.

Hydraulic actuators consist essentially of a double-acting piston housed in a metallic



Hydraulic Controller-Actuator

cylinder. Hydraulic fluid entering either one of the two actuator cylinder ports will cause the actuator piston to move. Actuator piston movement is transmitted through some appropriate mechanical linkage to the respective control device. (Short black arrows placed at various points in the controller, actuator, and transmission lines indicate the direction of flow.) Let us now consider each diagram individually.

Zero Error Signal

At A in the illustration, the components are in their neutral positions. For this condition to exist, the error signal input to the solenoid controller valve must be zero. You will notice that the valve is centered with respect to the various ports, and hydraulic fluid can neither enter nor leave the actuator unit. Therefore, both the actuator piston and the load to which it is linked are held in neutral positions.

Existing Error Signals

Assume first that the direction of error signal current flow through the solenoid has caused the controller valve to move to the extreme right as shown at B in the illustration. Fluid is thus permitted to pass through the controller unit into the right chamber of the actuator cylinder. The pressure forces the actuator piston toward the left. As the piston moves, hydraulic fluid in the left chamber of the actuator is displaced and returns to the reservoir as indicated by the arrows.

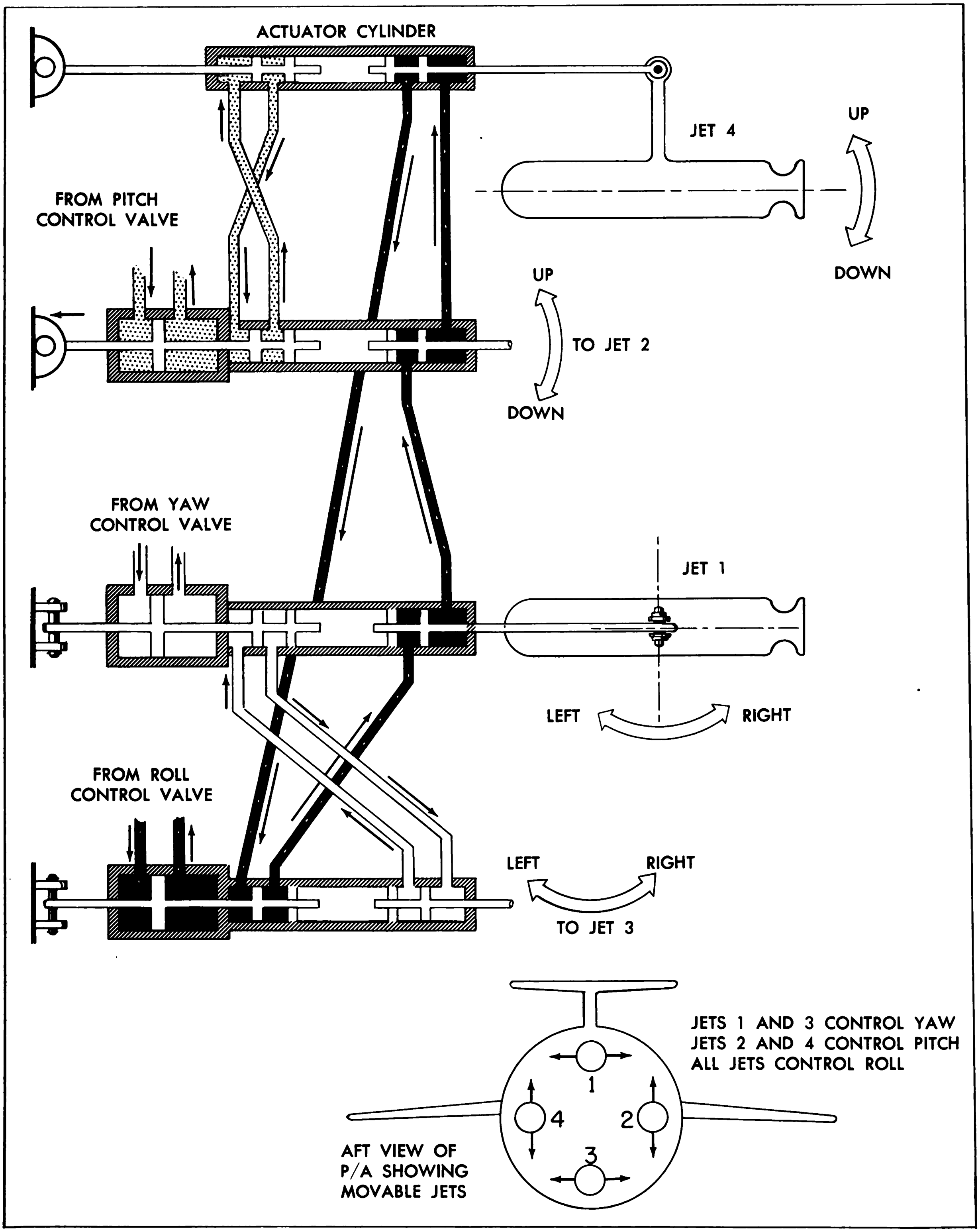
Now assume that an error signal exists which causes the solenoid operated controller valve to move to the left as indicated at C. Observe the direction of pressurized fluid flow through the controller unit to the actuator. Notice that the direction of piston movement is from the neutral position toward the right end of the actuator cylinder. The control device to which the actuator piston is connected will react accordingly and correct the attitude error. Note also the return path of the displaced hydraulic fluid from the right chamber of the actuator cylinder as it passes through the controller unit on its way to the reservoir. After normal flight attitude is resumed, all components will occupy neutral positions and remain so until the next attitude error occurs.

Any hydraulic system must be completely charged with fluid at all times to minimize the time lag between reception of an error signal and activation of the components. The *U-shaped tube* that runs from the outer left end of the controller housing to the outer right end allows for valve displacement and contains fluid when the valve is in its neutral position. This tube may be compared to a *shunt* in an electric system. However, when the valve moves to the right, as indicated at B, the fluid in the right end of the controller must be displaced or it will hinder the valve movement. The tube thus permits the trapped fluid to escape from the right end to the left end of the controller.

Certain characteristics of hydraulic transfer units make them highly desirable for use as controllers and actuators where outputs of from one to one hundred horsepower are required. For instance, you can see that the physical size of the moving parts is small compared to the large work output. Such units also lend themselves well to remote control by small electrical signals and to the transformation of electrical energy and the large amplification of mechanical power. Another advantage is the relative incompressibility of hydraulic fluids which permits rapid actuator response and accuracy in control displacement.

Hydraulic Interconnection of Control Channels

In the full page illustration you will find an interesting method used for controlling movable jets. Here the outputs of the three control valves (pitch, yaw, and roll) are used to position four jets. The required differential movement of the jets is accomplished by hydraulic interconnection of the four actuators. The method of obtaining hydraulic interconnection can be understood



Hydraulic Interconnection of Control Channels

by assuming displacement of each control valve separately and then following the resulting flow of hydraulic fluid through the cylinders to determine jet movement. From the information at the bottom of the illustration we find that jets *1* and *3* control yaw. The manner in which the hydraulic actuators are attached to the airframe and to the control arm on the jet indicates that they can only move to the right or left in a horizontal plane. However, upon investigating the installation of jets *2* and *4*, you will find that they can move only up or down in a vertical plane. Thus, jets *2* and *4* control pitch.

Now for a word about the actuators. This type (shown in the illustration) is known as a *floating cylinder*. Notice that the cylinders have a piston in each end. One piston is attached to the airframe and the other to the jet, thereby allowing the cylinder to *float* or move freely on the pistons in either direction. For example, if you trace the fluid flow from the pitch control valve, you will see that a pressure increase moves the *cylinder* to your left. The interconnection between jets *2* and *4* consists of an independent closed hydraulic system, having nothing to do with the fluid operating the valve. The tubing to the actuator of jet *4* is connected so that jet *4* moves in the same direction as jet *2* when an error signal is received through the pitch channel.

The interconnection of jets *1* and *3* for yaw control is similar to the pitch channel. Pressure from the yaw control valve moves the *cylinder* of jet *1* to your left, and the pressure transmitted to the actuator of jet *3* moves the *piston* connected to jet *3* in the same direction.

Differential movement of the jets required for roll control makes necessary the interconnection of all four actuators. Starting with the input signal to the actuator of jet *3*, we find that movement of the cylinder causes the jet to move to the *left*. The pressure transmitted to the actuator of jet *1* moves the piston so that the jet moves to the *right*. Pressure is next transmitted to the piston on jet *2* so that the jet moves *down*. Jet *4* will move *up* in response to the pressure change due to the action of jet *2*. Notice that *all four* jets can move clockwise or counterclockwise according to the sense of the signal in the roll channel. An advantage of this system permits corrections to take place in two or three channels at the same time. Thus, by simply changing the direction of thrust of the jets, an interceptor can be controlled about all three axes without the use of aerodynamic controls.

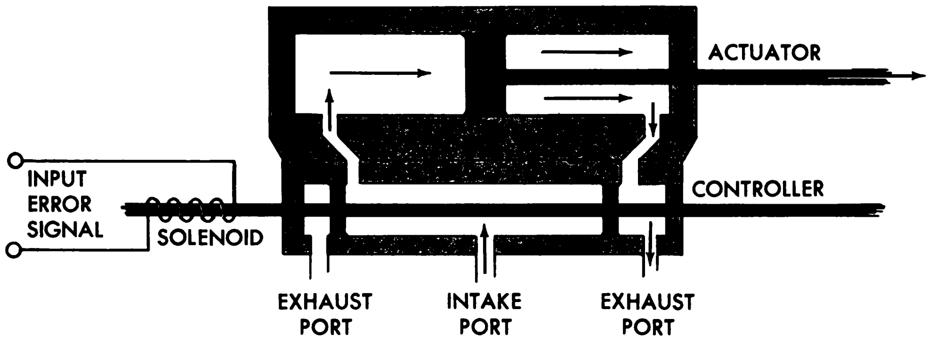
The above discussion presents only the basic principles of construction and operation of hydraulic controllers and actuators. Actually, numerous modifications of these units are used, depending upon the characteristics and requirements of the other control system components. However, any hydraulic controller-actuator unit will not deviate greatly from the construction and operation as described.

PNEUMATIC CONTROLLERS AND ACTUATORS

The construction and operation of pneumatic controllers and actuators are essentially the same as those of the hydraulic type. They differ in that a controlled flow of air under high pressure is used to displace the actuator piston instead of hydraulic fluid.

The figure illustrates the basic construction of a typical pneumatic controller and actuator. The controller unit consists of a solenoid-operated valve with an air intake port connected to an air supply, two exhaust ports, and two ports leading to the actuator unit. The pneumatic actuator consists of a double-acting piston which, when moved, transmits motion to a control device by means of a mechanical linkage.

Pneumatic controllers can be positioned by either a low-pressure air signal or by an electric error signal fed to a solenoid on the controller piston shaft. In either case, the sense and magnitude of the error signal are proportional to the attitude error of the craft involved.



Pneumatic Controller-Actuator

In reference to the illustration you will note that an electric error signal has caused current to flow through the solenoid of the controller unit. Notice that, as a result of the direction of current flow, the controller valve has been displaced to the left of its neutral position, and both ports leading to the actuator are open. Air under pressure enters the actuator chamber through the left port and forces the piston toward the right. This motion will be transmitted to the control device. At the same time, air on the right side of the actuator piston will be forced through the controller unit and exhausted.

Now suppose that the error signal and the resulting current flow through the solenoid cause the valve to be displaced in the opposite direction. In this case, the above action will be reversed.

Unlike a hydraulic system, a pneumatic system does not reuse its transfer medium after it has performed work on the controlled surfaces. Hence, air must be stored at a pressure much higher than that necessary for actuating the load so as to maintain adequate system pressure as the stored air supply diminishes.

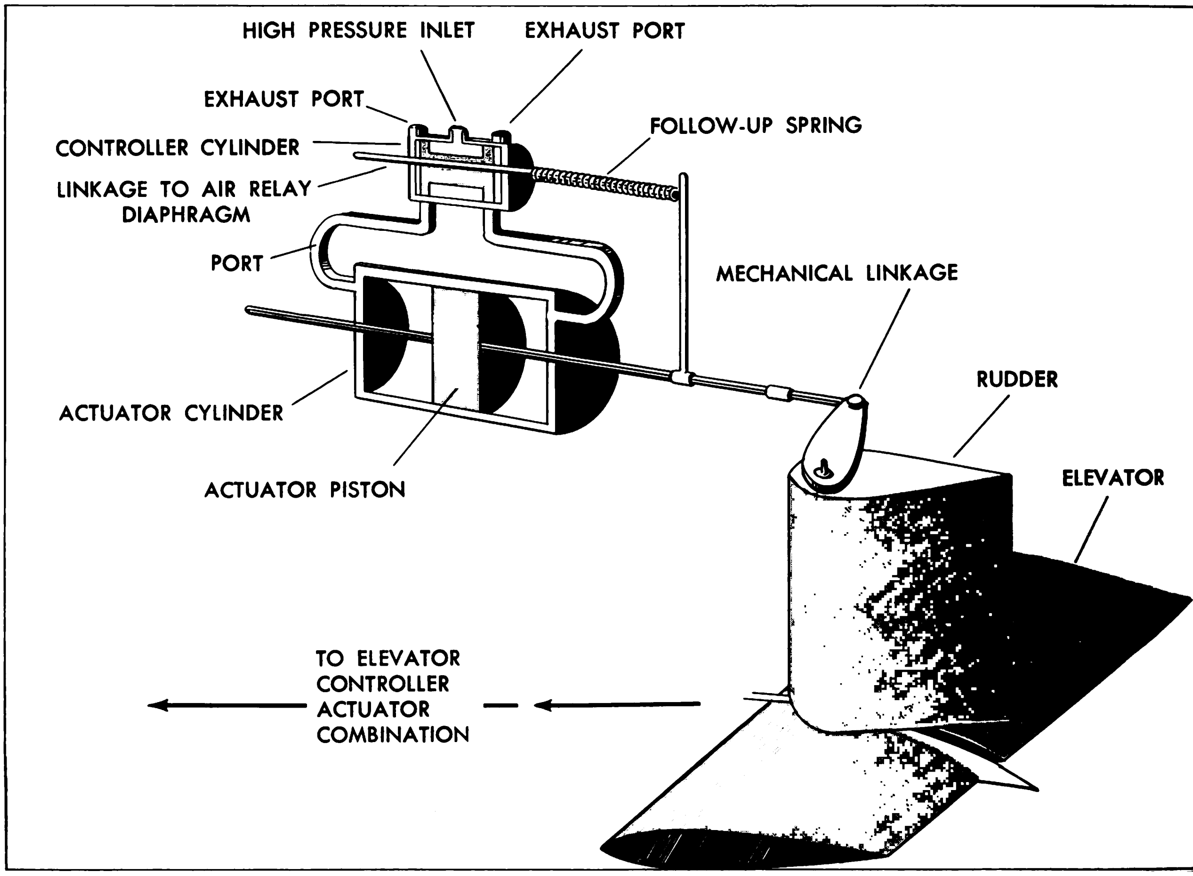
Rudder and Elevator Control

The diagram illustrates a double-acting piston-type pneumatic actuator used for rudder control.

The air control valve is mechanically linked to an air relay which receives an air error signal produced by azimuth deviations of the craft to be controlled. The air relay action is such that both magnitude and direction (sense) of the error signal are transmitted to the air control valve. Hence, the initial direction of displacement of the control valve is determined by the sense of the error signal.

Let us assume that an azimuth deviation has occurred and that the error signal from the air relay has caused the control valve to move toward the right, thus opening the right port to the actuator cylinder. Air from the high-pressure inlet passes through this port and causes the actuator piston rod to move toward the left-hand section of the actuator cylinder. This displaced air is exhausted into the atmosphere through the air control valve exhaust port located at the left end of the controller cylinder.

The resulting actuator piston motion is conveyed through mechanical linkage to the rudder, which applies corrective control in the proper direction to bring the craft back to its neutral position in azimuth. As the actuator piston moves, it exerts a force upon the followup spring. The followup spring is a calibrated coil spring connected between the actuator piston rod and the air controller piston rod. Hence, the followup spring, in



Pneumatic Actuating Units Mechanically Linked to Control Surfaces

turn, exerts a force upon the air control valve which opposes the force exerted by the air relay. In other words, the resultant movement of the air control valve is actually the difference of the two forces and is in the direction of the resultant force.

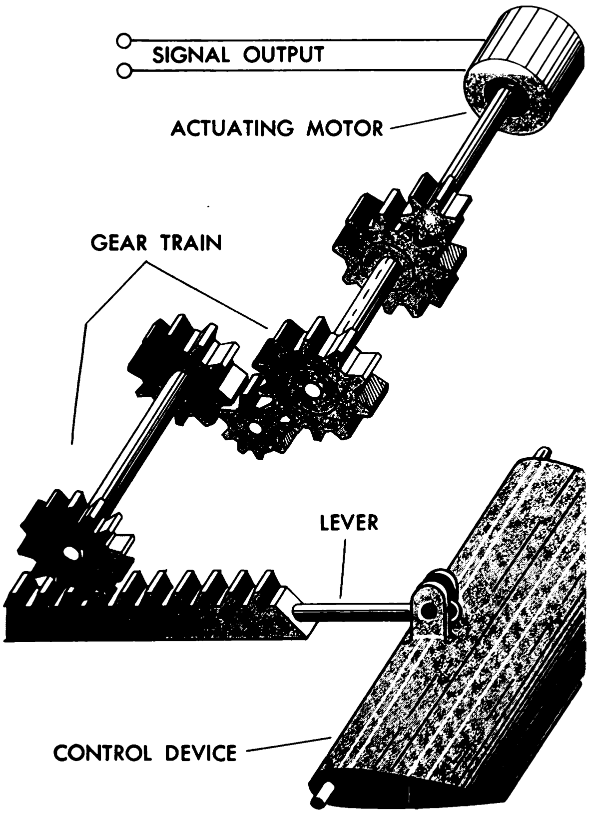
Movement of the actuator piston continues until the force exerted by the followup spring is equal, but opposite in direction, to the force exerted upon the control valve by the air relay. When this condition of equilibrium is established, the controller piston is centered, and movement of the actuator piston ceases. When the air control valve is thus balanced, air leaks past both sides of the controller piston and places equal pressure on both sides of the actuator piston. The above action holds the actuator piston and the rudder, to which it is linked, in the corrective positions commanded by the air error signal.

Electric Controller and Actuator Units

As you well know electrical energy can produce mechanical motion by means of magnetic force. Hence, either a solenoid or motor may be used as an electric actuator. However, a solenoid, normally, will not produce enough force to move an airfoil; therefore, motors are more often used.

It is not practical to apply the torque of a motor directly to the surface to be controlled. A motor used in this way would have to be of tremendous size in order to exert enough torque to move airfoils sufficiently. A large motor cannot be used, of course, because of its excessive weight.

A small motor running at high speed may have the same power potential as a larger motor that runs at some lower speed. Hence, a small motor is often connected to the control surface through a reduction gear train as shown in the accompanying illustration.



Gear Train Type of Mechanical Linkage

The mechanical advantage inherent in the gear train results in a large torque exerted on the control-surface pivot. The motor may be a constant-speed motor operating through a clutch, or it may be a variable-speed motor.

An Amplidyne Controller: DC Motor Actuator

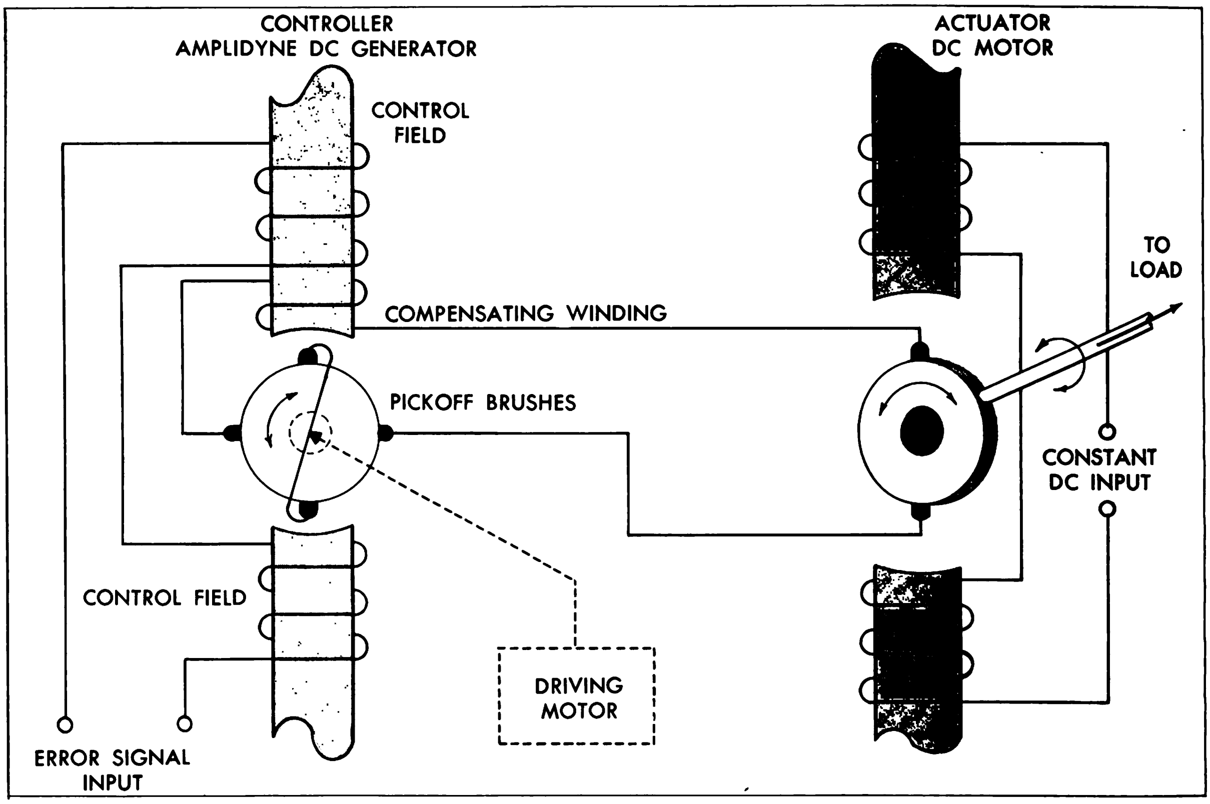
The schematic shown illustrates the manner in which a basic motor-generator combination (amplidyne controller) may be utilized to supply the proper energy to a DC motor actuator to activate a control device.

The controller consists of an amplidyne generator, the armature of which is driven by a constant-speed motor. The amplidyne may be considered as an amplifier, since a small amount of power applied to its control-field coil controls many times as much power at the output. An attitude error signal is used to excite the generator's control-field windings which are shown as a pair of field poles above and below the armature. Rotation of the armature within the control field causes a voltage to be induced in the armature coils. Since the armature coils are in a vertical plane when crossing the control field, the resultant induced voltage appears at the upper and lower brushes—numbered *1* and *2*.

If a load were to be connected to brushes *1* and *2*, the device would be an ordinary DC generator. However, in the amplidyne generator these two brushes are tied together or short-circuited. Hence, a resulting high (short-circuit) current flows through the armature. This current produces a second magnetic field that is much stronger than the initial control field and at right angles to it. The armature coils cutting this second magnetic field have induced in them a voltage which appears across the second pair of brushes—designated as *3* and *4*. This induced voltage is the output of the amplidyne.

This output voltage is proportional to the speed at which the amplidyne armature is rotated and to the magnetic field strength. Since the armature speed is constant, the output voltage varies with the magnetic field, and this, in turn, is proportional to the control-field excitation current or error signal.

As you can see, the actuator is a DC motor, the shaft of which may be mechanically linked through a reduction gear train to a control device. The field of the actuator motor is excited by a constant DC input, thereby maintaining a uniform magnetic field. Notice that the output of the amplidyne generator is fed to the armature



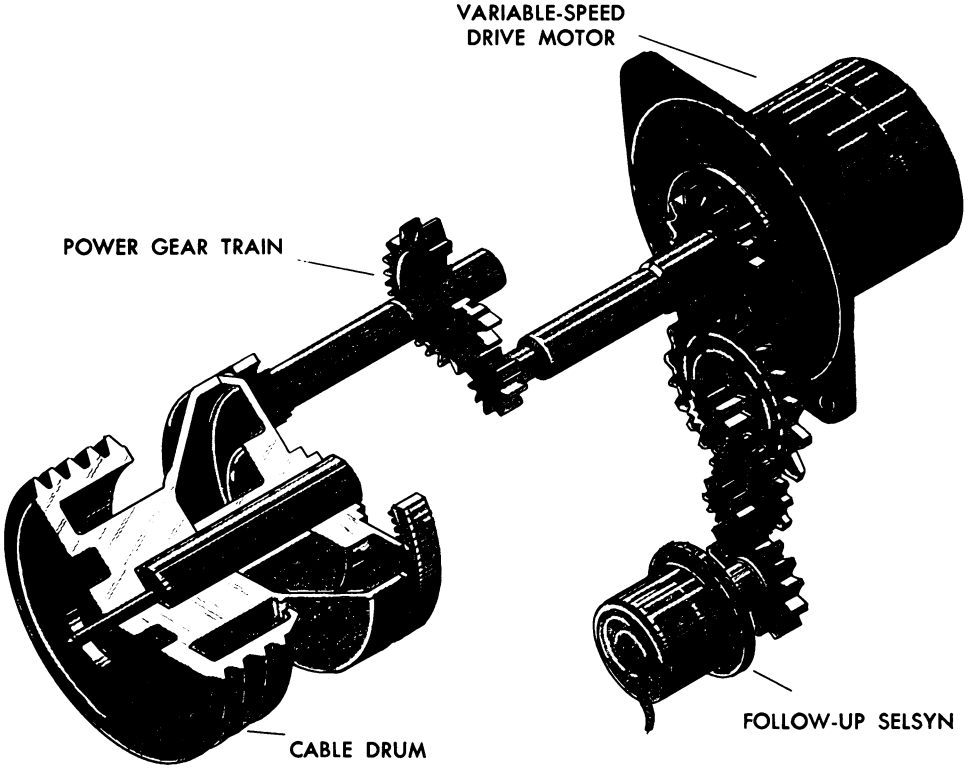
Electric Control-Actuator

brushes of the DC motor. The resulting current flow through the DC motor armature coils produces a magnetic field, which reacts with the polar field and causes the armature to rotate. Since the polar field is constant, the magnitude and direction of armature rotation are determined by the magnitude and direction of current flowing through the armature coils. This, in turn, is proportional to the magnitude and polarity of the error signal input to the control field of the amplidyne generator.

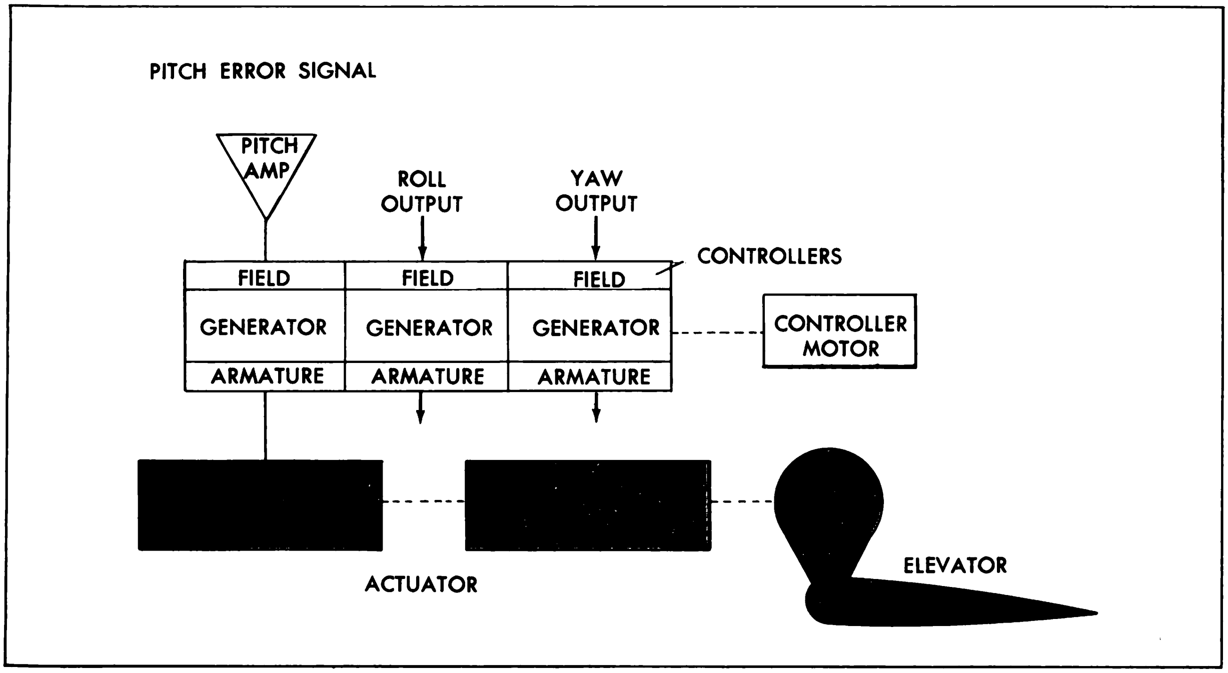
The compensation winding shown on the upper field pole of the amplidyne is used to eliminate one of the major difficulties that arises in this type of controller unit. In the absence of the compensating winding, the load current flows through the amplidyne armature coils, producing a magnetic field which tends to cancel the control field. This cancellation tendency is an undesirable feature because it reduces amplification and causes poor regulation of the amplidyne system. In order to compensate for this opposition to the control field, the compensating winding is added to the field structure. The magnetizing force produced by the passage of load current through the compensating coils can be made to balance that produced by the flow of output current through the armature coils.

Variable Speed Actuator: Constant-Speed Controller

The figures on the following page illustrate the use of a variable speed motor to actuate a control surface. Here again, the motor rotor rotates in either direction, depending on the sense of the error signal. Furthermore, the rotor will accelerate rapidly and



Variable-Speed Electrical Actuator



Part of an Electrical Attitude Control System

will revolve at a speed roughly proportional to the strength of the error signal. The motor rotor is coupled to a control surface through a reduction gear train; therefore, the control surface movement will be proportional to the speed of the motor rotor.

A functional block diagram of portions of an electrical attitude control system is shown accompanying the variable-speed actuator. The controller, shown as part of this system, converts the power of the controller drive motor to the three variable-speed channel motors of the type illustrated in the accompanying diagram. The controller must be capable of producing large output power; that is, the signals fed to the variable-speed motors must be large because it is these signals that supply the driving power. In some systems the variable-speed motors are driven directly by the output of an electronic power amplifier. However, in the system depicted here, the actuators are driven by the output of their respective generator-type amplifiers which themselves receive the amplified version of the error signal.

For example, suppose a pitch error signal is received and is amplified by the pitch amplifier. This increases the magnetic field of the pitch generator. The voltage output of the pitch generator is, in turn, increased, and the required amount of power is fed to the variable-speed actuator motor.

It should be pointed out that the controller drive motor must maintain a reasonably constant speed regardless of load. Otherwise, if the pitch output decreased the speed of the controller motor and generators, the output of the roll and yaw generators, would also be decreased at that instant. This controller speed variation would cause undesirable cross-coupling between channels since the pitch signal would affect the other channels or vice versa.

Constant-Speed Actuator

The inertia effects of starting and stopping a variable-speed motor can be eliminated by using an actuator motor which runs continuously and maintains rather uniform speed. In such a case the motor is connected to a control surface through a clutch which serves as the controller for the system. The clutch varies the amount of power transmitted from the motor to the control surface. The use of two clutches and a gear differential will allow control in both directions. The friction clutch discs are made to contact each other by means of solenoids controlled by a channel power amplifier; and the amplifier needs to supply power only to operate the solenoids. Such a clutch system is used in the autopilot discussed in the next section of this chapter.

ADVANTAGES AND DISADVANTAGES OF ACTUATOR MOTORS. The required fast rotation of a small motor introduces a major disadvantage in an electrical system. Suppose a sudden deviation of an interceptor should require rapid action from a control surface. This means that the rotor of a small motor must revolve at a high speed to produce the required rapid response. A large amount of inertia exists when suddenly trying to reach a high rotor and gear-train speed from a standstill. This resultant inertia opposes any required velocity change and causes an undesirable lag in the response of control surfaces. The lag may be so great that the system either operates with insufficient sensitivity or with a tendency to oscillate.

Even if a large motor could be used, the motor would still possess inertia due to its large mass. One means of partially compensating for this disadvantage is to control the output of a continuously operating motor by means of a clutch. A further improvement has been in the development of small, high-torque motors.

The major advantage of an electric system is that the system uses only one medium—*electricity*—for operation. This reduces supply, assembly, personnel, and shipping problems.

SECTION C, ELECTRONIC ATTITUDE CONTROL

A major advantage of an electronic attitude control or space stabilization system is that its operation requires only one medium (electricity) which is conveniently transmitted through electrical wiring. It is also light in weight and comparatively trouble-free.

Such a system is fast in response, accurate, and easily tied in with radio and radar guidance systems. It is affected less by extreme temperature changes. An electronic system is also more difficult to disable by enemy action.

The heart of electronic attitude control systems is the widely used, electrically-powered automatic pilot.

AUTOPILOT

An autopilot is an electromechanical device which can automatically fly an aircraft in straight and level flight or maneuver it in response to fingertip controls operated by a human pilot. An autopilot may maintain an interceptor on an accurate course by receiving directional control or command intelligence directly from a radar system. An autopilot may be designed to keep an interceptor in a predetermined attitude and on a selected heading for long periods of time with a minimum of attention. When used in this manner, directional control is accomplished by coupling the autopilot directly to a slaved gyro-magnetic compass system or to a gyro-directional stabilizer.

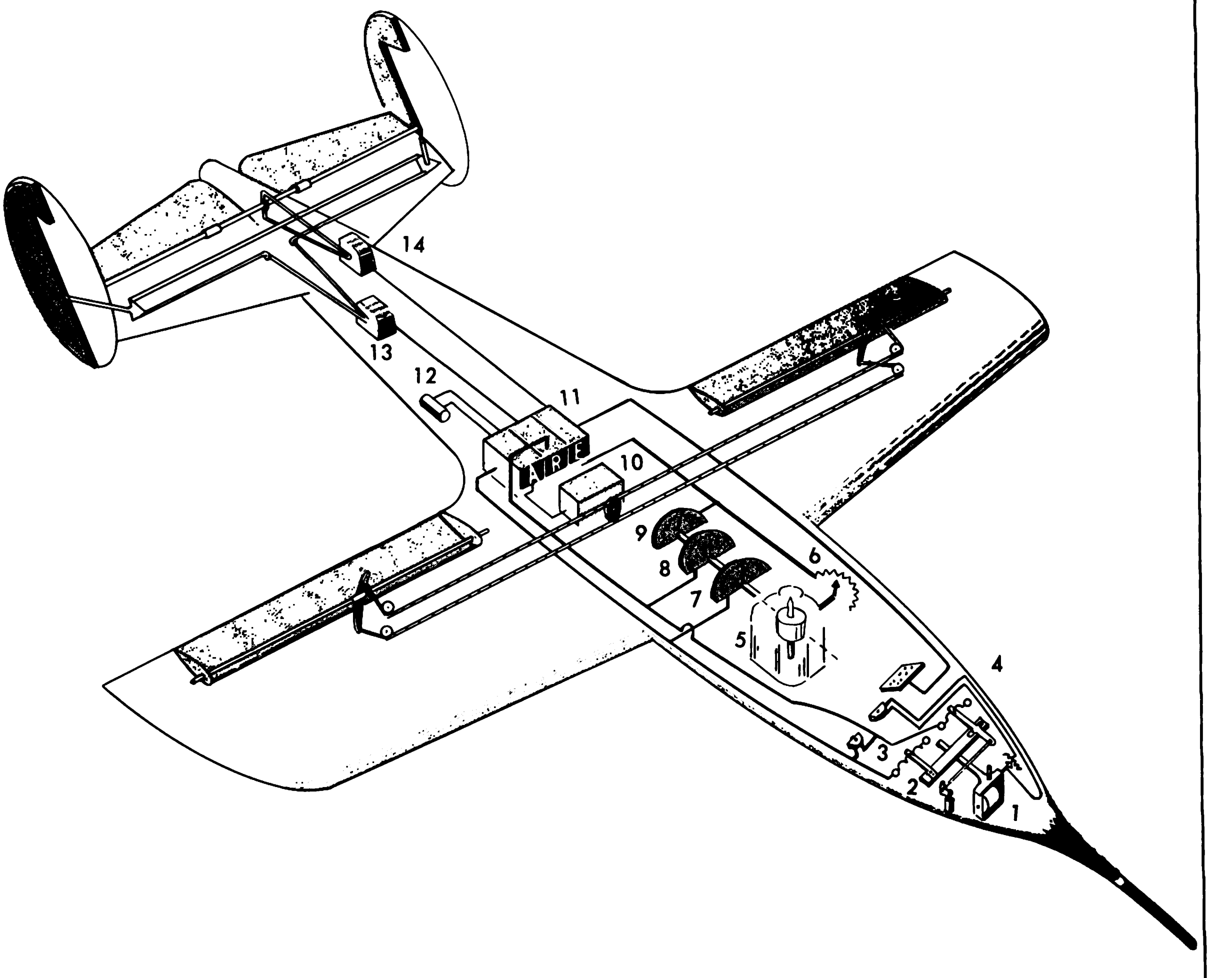
A typical autopilot consists of various separate units electrically interconnected to operate as a system. A general overall understanding of their functions and relation to each other can be gained by referring to the full-page illustration. Your attention is called to the fact that in this illustration the connecting lines do not represent wires but serve to connect related units.

Typical Operation of an Autopilot

Let's assume that the autopilot is in operation and that the interceptor in the illustration is flying straight and level. Now if the interceptor turns away from its established position, the gyro-operated directional stabilizer (*1*) will detect or sense this deviation and move the potentiometer wipers within the directional panel (*2*) to one side or the other. The banking potentiometer (*3*) and rudder pickoff potentiometer (*4*), located in the directional panel, are electrical devices which send signals to the aileron and rudder section of the amplifier (*11*) whenever a signal is detected by the directional panel (*2*). The amplifier signals are then converted, by means of magnetic switches (relays), into electrical impulses which are applied to the aileron and rudder actuators (*10* and *13*), causing them to operate the ailerons and rudder in the proper direction and amount to bring the interceptor back to its original position.

Upon further investigation of the illustration, you will find that, if the nose of the craft should drop, the vertical flight gyro (*5*) detects the resulting vertical deviation and operates the elevator pickoff potentiometer (*6*) which feeds an electrical signal to the elevator section of the amplifier (*11*). Here the signal is amplified and relayed in the form of electrical impulses to the elevator actuator (*14*) which, in turn, raises the elevators the proper amount to bring the interceptor back to level flight.

If one wing should drop appreciably, the vertical flight gyro (*5*) operates the aileron pickoff potentiometer (*7*), the skid potentiometer (*8*), and the up-elevator potentiometer (*9*). The electrical signals, caused by the operation of these units, are transmitted to their respective (aileron, rudder, and elevator) sections of the amplifier. The resulting impulses to the aileron, rudder, and elevator actuators cause each of these units



1 DIRECTIONAL STABILIZER
2 DIRECTIONAL PANEL
3 BANKING POT
4 RUDDER PICKOFF POT
5 VERTICAL FLIGHT GYRO
6 ELEVATOR PICKOFF POT
7 AILERON PICKOFF POT
8 SKID POT
9 UP-ELEVATOR POT
10 AILERON SERVO
11 AMPLIFIER
12 ROTARY INVERTER
13 RUDDER SERVO
14 ELEVATOR SERVO

Stabilizing Components of a Typical Autopilot

to operate its respective control surface just enough to bank and turn the interceptor back to a level flight position. This is an excellent example of how the three channels are operated simultaneously in order to perform a coordinated turn.

When performing such a turn, the control sends signals through the aileron and rudder sections of the amplifier. The signals are also sent to the aileron and rudder actuators which operate ailerons and rudder by the proper amount. This completes a perfectly coordinated (nonslipping, nonskidding) turn. As the interceptor banks, the vertical flight gyro (*5*) operates the aileron, skid, and up-elevator potentiometers (*7*, *8*, and *9*). The resulting signals from the aileron and skid potentiometers cancel the signals to the aileron and rudder actuators in order to streamline these controls during the turn.

The signal from the up-elevator potentiometer causes the elevators to rise just enough to maintain altitude. The up-elevator potentiometer is center-tapped in order to insure up-elevator movement in either a right or left turn. When the desired turn is completed, a reverse signal is applied and the interceptor levels off to its new heading. It is important for you to know that a switch energizes a device known as a *directional arm lock* on the directional stabilizer. This device prevents the stabilizer from interfering by performing its normal *direction-correcting* function during the turn.

The rotary inverter (*12*) is a motor-generator unit which converts direct current from the battery into 105-cycle alternating current for the operation of the autopilot.

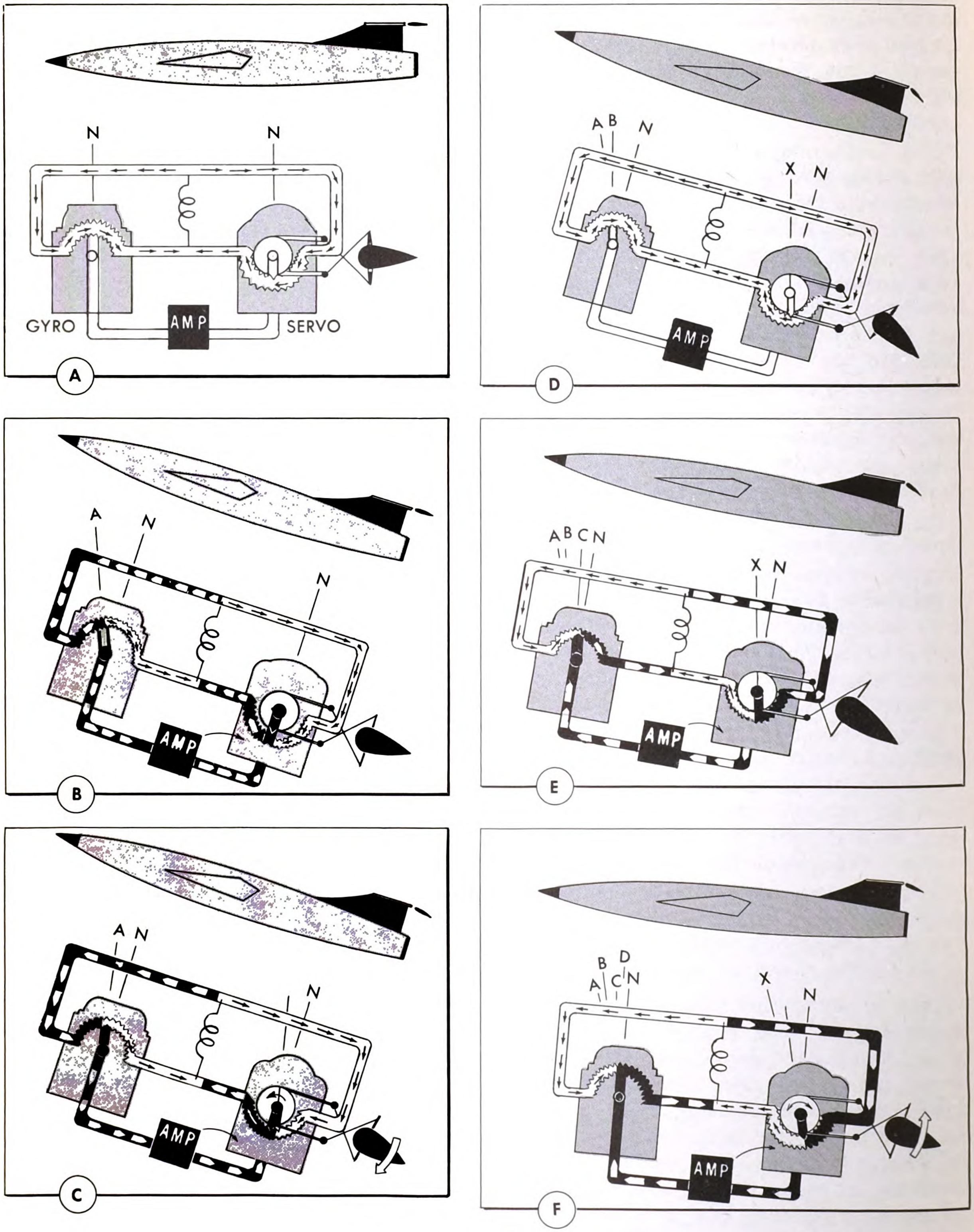
PITCH STABILIZATION. The next series of illustrations and their accompanying explanations describe the step-by-step interaction between the vertical flight gyro and the amplifier and actuator resulting from a deviation of an interceptor in its pitch axis only. The autopilot operates in a similar way to correct deviations in the roll and yaw axes. The diagrams illustrate how an interceptor's deviations cause the gyro to unbalance a bridge circuit and how the servo unit rebalances the circuit by following the gyro movements. This is the fundamental principle involved in the operation of a typical automatic pilot.

As depicted at **A** in the series of illustrations portraying how an autopilot corrects for pitch deviations, an interceptor is assumed to be in straight and level flight. Hence, the elevator pickoff potentiometer in the gyro is at its normal position (*N*) and the servo balance potentiometer is centered, holding the elevator control surfaces in their normal, level-flight position. Under this condition the bridge circuit is balanced, and no correcting signal is applied to the amplifier.

Now assume that a vertical gust of air, as shown at **B**, raises the nose, causing a deviation in the pitch axis. Notice that the gyro case and its attached elevator pickoff potentiometer tilt with the interceptor causing displacement of the potentiometer wiper to position *A* with respect to the potentiometer. This movement unbalances the bridge circuit resulting in the application of a correcting signal to the elevator channel of the amplifier. This causes the *down-elevator* relay in the amplifier to close, sending a *down-elevator* impulse to the elevator servo unit.

At **C** in the series, we find that the servo unit responds by turning counterclockwise, driving the elevator control surfaces downward by means of cables connected as shown. Note that the direction of actuator rotation is such that the balance potentiometer wiper approaches a position where the bridge circuit will again be balanced. The bridge circuit is balanced only when the balance potentiometer wiper and the pickoff potentiometer wiper are so positioned that the signal voltage is zero.

As soon as the servo unit begins to apply *down-elevator* as depicted at **D**, the interceptor begins to level off and the pickoff potentiometer in the vertical flight gyro begins to return to its normal position. At some intermediate position (*B*), the bridge circuit



How an Autopilot Corrects for Pitch

will again be balanced as the balance potentiometer wiper has by this time moved to position *X*, and the wipers are again positioned so that the signal voltage is zero. At this point, servo rotation will stop, as sufficient down-elevator has been applied to produce the required correction.

It is now necessary to reverse the direction of control surface movement in order to prevent *over control.* Figure **E** in the series will illustrate how this is done. As the interceptor continues its return to level flight, the elevator potentiometer moves to position C, resulting in an unequal voltage on the balance potentiometer. Thus, the bridge circuit is again unbalanced, *but this time in the opposite direction.* The effect of this unbalance on the amplifier causes the up-elevator relay to close, resulting in clockwise rotation of the actuator and up-elevator movement.

As the actuator rotates clockwise, returning the elevator control surface toward its normal flight position, the interceptor continues its return to level flight. However, since the control surface displacement is being gradually reduced by a follow-up signal, the speed with which the interceptor returns to level flight is gradually and smoothly diminished. Thus, the clockwise movement of the pickoff potentiometer becomes slower and slower as the interceptor approaches level flight—see **F** of the series. This action continues until the interceptor reaches its normal level-flight attitude, at which position the balance potentiometer catches up to the pickoff potentiometer and the bridge circuit is again balanced with both potentiometers electrically centered. You can see that the separate steps described above actually take place in a gradual progression, each movement blending into the next to produce smooth correction which may be rapid at first, but which tapers off smoothly as the interceptor resumes its normal flight attitude.

DESCRIPTION OF COMPLEMENTARY ELECTRICAL COMPONENTS. The heading reference is established by an electrically driven gyro and sensed by a selsyn or reluctance pickoff. Any computing operation on the error signal is by electrical means. A rate signal would be obtained from the output of an electrically driven rate gyro or from an electronic rate circuit operating on the displacement gyro signal. Similarly, an integral signal would be obtained by a motor driven selsyn or by an electronic integrator circuit. Voltage amplification would be obtained by electronic means or by using a dynamotor type amplifier such as a motor-generator set or an amplidyne.

SECTION D, INERTIAL NAVIGATION AND GUIDANCE

An inertial navigation and guidance system is looked upon as the truly ultimate of *preset* guidance systems, because the system may be independent of any outside source of information. An interceptor aircraft or missile, once launched and utilizing inertial guidance, would find its way to a predetermined target location without a single visual or radio data link outside of its airframe.

Any guidance system which does not depend on outside information is highly desirable because it cannot be *jammed.* Jamming, you will remember, is enemy interference of guidance information in order to prevent an interceptor or missile from reaching its target. The only countermeasure against inertial guidance is destruction of the guided craft itself, and the probability of such a countermeasure can be reduced by using supersonic, high-altitude interceptors and missiles.

In this section we will discuss the fundamental principles and investigate some of the basic components incorporated in inertial

guidance and navigation systems. The study of inertial concepts will show you how a craft's accelerations are detected and how these accelerations may be used for guidance.

An inertial guidance system makes use of Newton's second law of motion which states:

> **An unbalanced force acting on a body causes the body to accelerate in the direction of the force, and the acceleration is directly proportional to the unbalanced force and inversely proportional to the mass of the body.**

Remember, acceleration is the rate at which a body is changing velocity. The law also applies to deceleration, which is the rate at which velocity is decreasing. The force due to acceleration is the same force that presses you against an auto seat as the auto picks up speed. It is also a similar force in the opposite direction (deceleration) that throws you toward the windshield when the auto suddenly slows down. Thus, acceleration of a body is proportional to the outside force exerted upon the body.

In an inertial system, acceleration may be measured by means of an inertial force of acceleration upon a mass. Since this acclera-tion has an effect on flight, an inertial guidance system changes the flight path to compensate for undesirable changes in accelerations.

Inertial guidance may be used on both short and long-range craft. Regardless of the trajectory or flight path, complete control by inertial guidance requires detection of accelerations along the three axes of movement. Movement to right or left along the pitch axis is known as *lateral motion,* movement forward or aft along the roll axis is called *longitudinal motion,* and movement along the yaw axis produces an *altitude change*. Note that the term *along* an axis rather than *about* or *around* an axis is used. You must remember, along an axis means *parallel* to the axis. This is very important in understanding the concepts of inertial guidance.

INERTIAL SYSTEM OPERATION

The primary sensing device used in an inertial system is an *accelerometer*. This instrument is capable of detecting any straight line displacement deviation. Such motion is known as *translation*. In this case, we are *not* interested in attitude deviations. In an actual system, acceleration may not always be detected along all three axes. Instead of an accelerometer, some other sensing device such as a gyro or an altimeter may be used.

You will recall that the altitude, range, and azimuth are some of the factors than can be preset prior to launching. These same three factors must be preset into an inertial system. When launching a craft using only an inertial system, two facts must be known: first, the *exact distance* between a starting point and the target; and second, the *exact direction* from the starting point to the target.

In making azimuth settings, allowance is usually made for winds. This compensation would be made for an average crosswind. As you know, however, wind cannot be accurately predicted and at present there is no means of accurately detecting the amount of drift by means of a gyro or an airspeed device.

In an inertial system, course *direction* may be set in by using a gyro for azimuth reference. *Side drift* could be detected by an accelerometer oriented to detect lateral acceleration. This acceleration will then be fed to a computer which will compute the *distance* to the right or left of course. This distance error signal is then used to turn the craft back to the desired heading by feeding the signal into a yaw control channel as will be shown presently. It might be well to first investigate the inertia device—the accelerometer—used to detect the degree and direction of a trajectory change.

Accelerometers

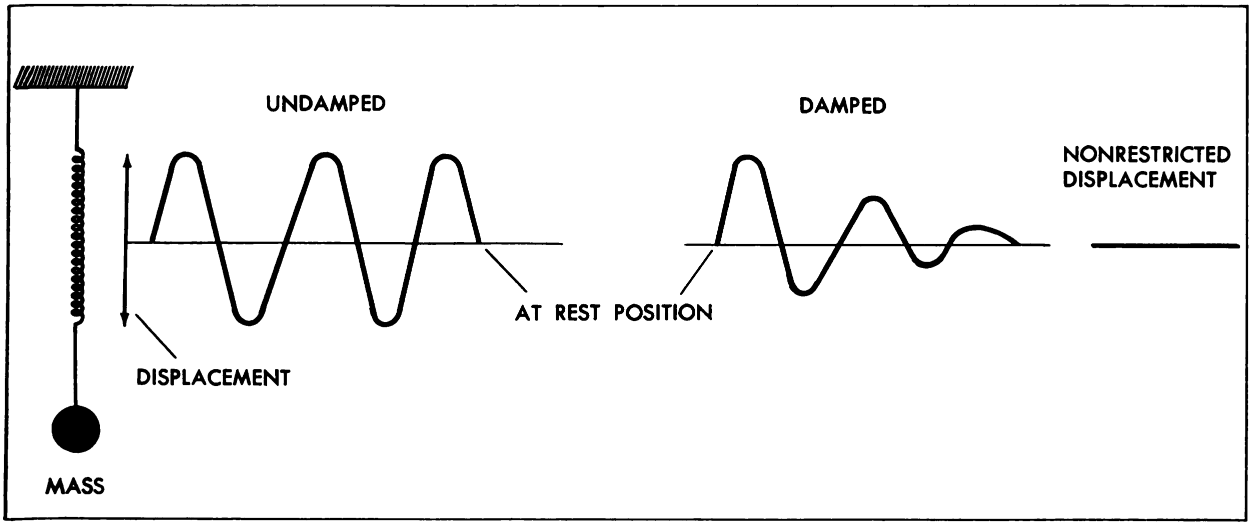
In the current development of accelerometers, in order to obtain more accurate units, there are completely different approaches to the problem. Nevertheless, every type of accelerometer has certain common features. Each accelerometer has within its case a mass which is free to move between certain limits. This mass, sometimes called the converter, is the part upon which the acceleration force acts. This mass converts acceleration into a measurable quantity.

We find that all accelerometers operate on one of two principles. Using one principle, the inertial reaction force causes a displacement of the converter in an elastic mounting system (for example, a set of springs). The greater the force, the further the displacement. This displacement is then measured by several methods. The second type operates on a fundamentally different principle in which the force that counteracts the inertial reaction force of the mass is not supplied by an elastic mount, but is supplied by an electric current, a stream of air, or any other medium which can produce a controllable force. In this system, a very small deflection of the mass is detected, and a counteracting force is instantly applied to the mass to prevent any further motion. The acceleration is then indicated by the magnitude of this counteracting force.

Accelerometer Damping

Whenever a mass, supported by an elastic mounting system such as a set of springs, is displaced due to some force, it will tend to oscillate about its initial—at rest—position. This oscillation is known as *simple harmonic motion* and is the same as the oscillation of a pendulum. (A pendulum is a simple example of an accelerometer.) Damping reduces the amplitude of each oscillation, thereby limiting the number of oscillations. The natural damping action may be increased by the use of a device designed especially for this purpose.

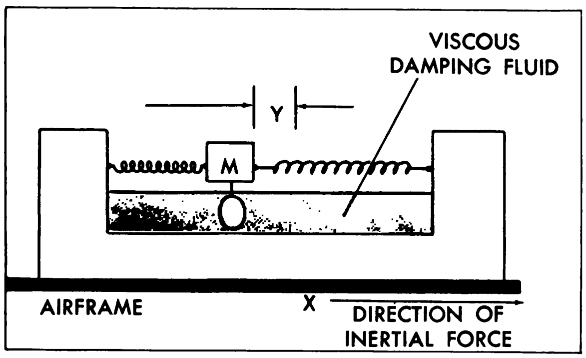
For example, if you were to displace the spring-suspended mass (shown at the left in the illustration) and then release it, it would tend to oscillate in simple harmonic motion. If there were no retarding action on the spring, it would oscillate as illustrated in the *undamped* condition. However, a spring does offer resistance to the oscillation, causing it to diminish as shown in the slightly damped condition. In order to have an efficient accelerometer, it is necessary that the



Effect of Damping Oscillations

amount of damping be just great enough to give a nonrestricted displacement and yet prevent any oscillations from existing as illustrated in the third condition.

Another type of damping by means of an oil reservoir and paddle is illustrated in the accompanying diagram. In this case, suppose we assume that a missile is to accelerate suddenly by an amount Y. The mass M would also be displaced (relative to the missile) by some distance in terms of X. Furthermore the relative displacement of the mass M during this time would indicate a force present—remembering that $F = Ma$ where $F =$ force, $M =$ mass, and $a =$ acceleration.



Displacement of Viscous-Damped Accelerometer Indicates Acceleration

There will be a certain displacement of the spring per unit of force applied. The spring at the end of this period of time will tend to reverse the direction of force on the mass. This would lead to simple harmonic motion. The damping is just sufficient to eliminate this oscillation. Therefore, we have a system that will be sensitive and at the same time tend to prevent any transient or unwanted oscillations. The accuracy of such a system depends upon the method used to measure the displacement of the mass relative to the missile and on the linearity of the elastic mounting. The sensitivity of the system may be increased by reducing the frictional forces in the mounting and by improving the displacement pickoff.

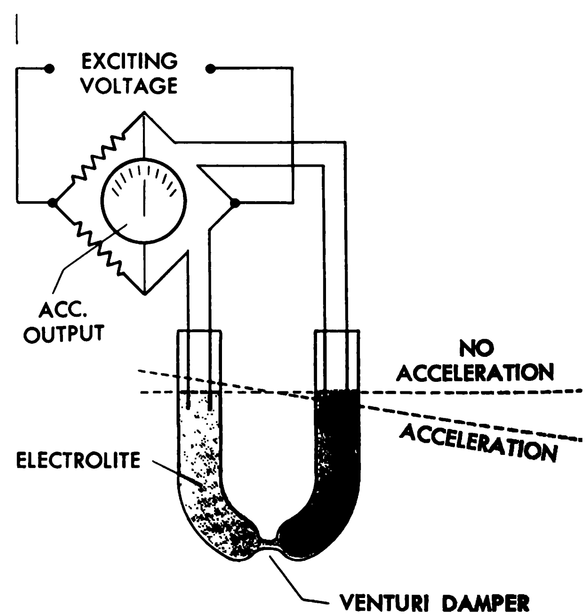
Types of Accelerometers

There are many ways of obtaining an accelerometer signal which is a function of the displacement of a mass. For example, one type developed by M.I.T. utilizes a change of inductance between two coils as their separation is varied. The coils form part of an inductance bridge and, as the inductance changes, the bridge is unbalanced and an AC voltage output results.

Another accelerometer uses wire strain gages as the suspension elements for the mass. The strain gages form the arms of a bridge. A change of acceleration causes a change of the electrical resistance of the circuit, giving an AC output that is an indication of the acceleration.

Accelerometers are also used in which the capacitance of a circuit is varied proportionally to the acceleration. This capacitance forms part of a capacitance bridge circuit which becomes unbalanced when acceleration or deceleration occurs.

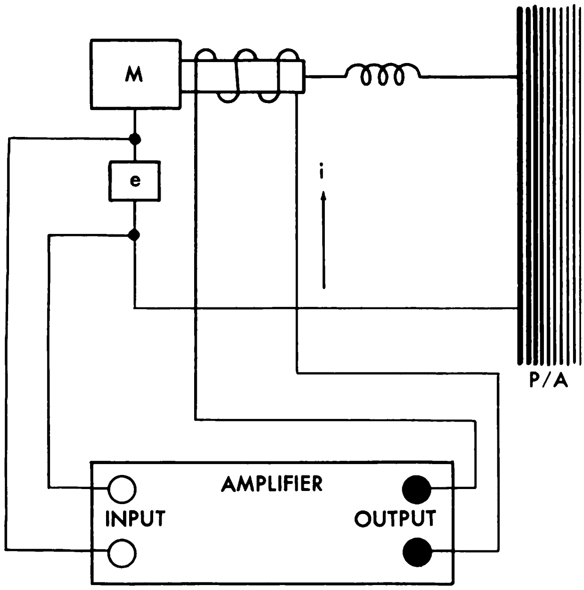
Another type of accelerometer is the manometer as shown in the accompanying diagram. Accelerations are indicated by the



Manometer Accelerometer with Resistance Bridge Pickoff

electrolyte flowing towards one or the other end of the manometer, thereby varying the resistance between pairs of electrodes. The venturi dampens the response of the manometer by limiting the electrolyte movement. This damping prevents oscillation of the electrolyte, thereby preventing oscillation of currents fed to the computers.

Shown next is an accelerometer which operates on the principle that the force which counteracts the inertial reaction force of a mass is not supplied by an elastic mount, but by an electric current.



Accelerometer in Which Inertia Reaction Force of Mass is Supplied by an Electric Current

For example, let e be a voltage that is developed due to the displacement of the mass, and i, the output of the amplifier (the restoring current fed to the coil). If e is dependent on the displacement of the accelerometer mass M, then the output of the amplifier will be dependent upon the displacement. For example, if the craft to be controlled should accelerate, there will be a certain voltage applied to the amplifier input. After necessary amplification, there will be a proportional output applied to the coil in such a manner as to counteract almost all of the inertial force. The acceleration, therefore, could be measured by the displacement of the mass, by the resultant input to the amplifier, or by a restoring force resulting from the current fed to the coil.

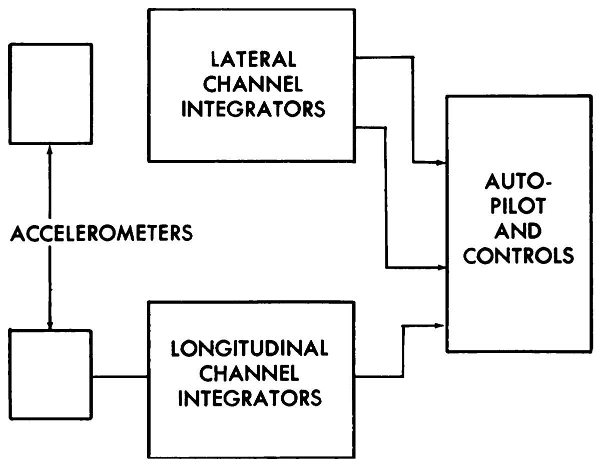
The amount of displacement allowed by the mass is extremely small. This means that errors which may arise due to the nonlinearity of the electric field are mostly eliminated. The sensitivity of such a system may be increased with little effect on the range of operation.

There are many ways of measuring acceleration. The methods just described are samples of the many possible ways used to accomplish this function in an inertial system.

Inertial System Computer

Acceleration data as such cannot be used for guidance; distance data must be used. A functional block diagram of a basic inertial computer is shown in the accompanying illustration.

The system illustrated detects errors in the desired flight path by measuring sideward (lateral) accelerations as well as fore and aft (longitudinal) accelerations during flight.



Inertial Guidance Computer

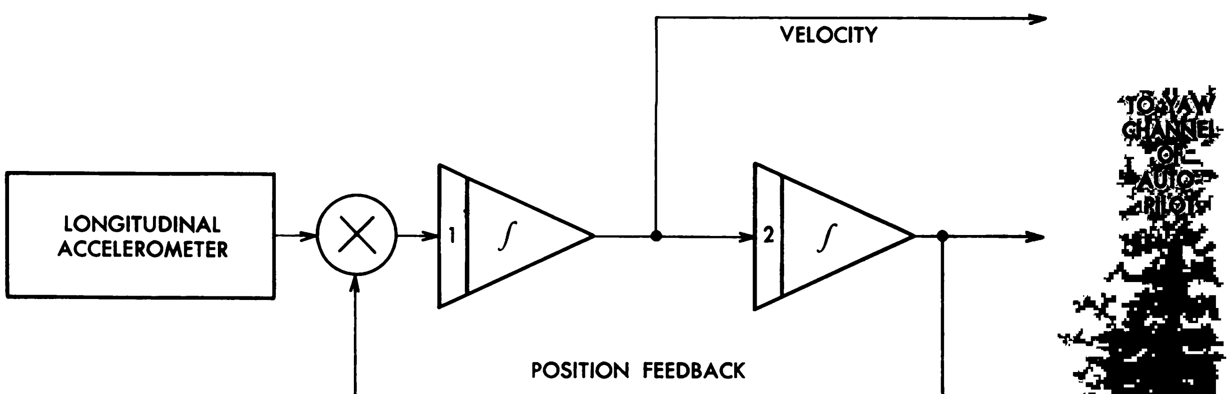
The computer equipment consists of two main channels: the lateral channel and the longitudinal channel. The channels are quite similar in arrangement. Each channel contains an accelerometer and a method of double integration. The accelerometers detect velocity changes without the need of a reference that is exterior to the airframe. The components which change the detected accelerations to distance data in each channel perform the mathematical function known as double integration. In order to do this, two integrators are connected in series. The input to the first integrator is the accelerometer signal. This signal is integrated to produce an output which represents the *velocity* which would result from the acceleration. This velocity signal is then fed to the second integrator which produces a signal indicating the *distance* which would be produced by the velocity input. The final result, then, is a signal which represents distance traveled along a certain axis as a result of detection by the accelerometer.

Integrator components of computers were discussed in the preceding chapter. Although there are many different types, they all tend to produce the same output from a given signal. The main difference is in the means of producing this output and in the varying degree of accuracy of various types. You will recall that the application of integrators in a control system is to produce a signal which represents an accumulation of a particular error. In an inertial guidance system, velocity results because mathematically it happens to be the first integral of acceleration. Distance results because it is the second integral of acceleration.

LATERAL CHANNEL. A breakdown of the lateral channel may appear as shown in the accompanying diagram. (Notice that the standard symbol for integrators is used.) Remember that the lateral channel produces an output that represents the displacement distance which is a result of sideward acceleration to the right or left of the desired course.

The lateral channel output should be zero at all times if the guided craft is on course. If the craft drifts off course, the computer output (second integration) is a voltage whose magnitude and polarity are dependent on the distance and direction of off-course displacement. The first integrator of the lateral channel will have an output voltage proportional to the velocity of motion away from or toward the desired course. This is a guidance rate signal. Both of these voltages may be utilized in an autopilot to determine the amount and direction of control to be applied in a transverse direction to the course. These signals may be necessary because the autopilot may detect heading (attitude) errors, but not off-course conditions.

Let us now assume that a crosswind begins to blow a missile off to the right of its predetermined course as depicted at A in the



Lateral Channel

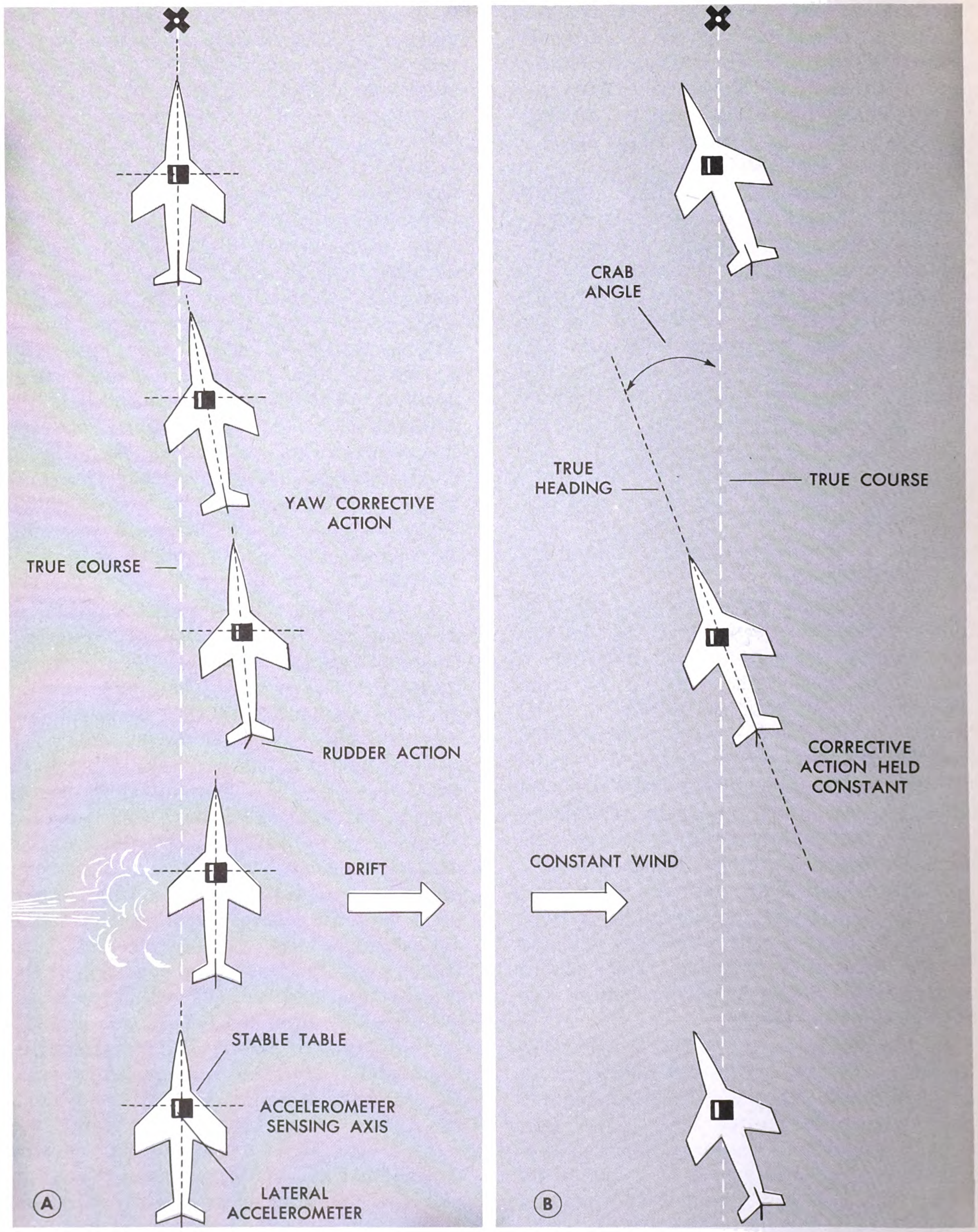
next illustration. This wind will create a sideward acceleration which immediately starts the integrators computing the distance error. Although the acceleration may have completely stopped after a moment, the missile has acquired a velocity to the right of its course as illustrated. This velocity is indicated by the output of the first integrator (velocity integrator). The signal is then fed to the second integrator (distance integrator) which continues to build up a signal to represent distance traveled to the right due to the sideward velocity. As the yaw control channel (of perhaps an autopilot) reacts to the distance error signal by producing a nose left attitude, the sideward velocity to the right decreases. This is the same as saying the missile decelerates. This deceleration affects the first integrator whose output decreases until it indicates zero sideward velocity. Thus, the distance error recorded on the second integrator output will increase no further.

It is now necessary for the opposite action to take place in order to return the distance error signal from the second integrator to zero. (This also means reducing the actual sideward deviation of the missile to zero.) As the yaw control system moves the missile back to the desired course, acceleration to the left produces a velocity integrator output of opposite sense which eventually reduces the distance output to zero. At this time the missile is again on its desired course. Bear in mind that a stable table is required to maintain the lateral accelerometer perpendicular to the true course, not perpendicular to the interceptor heading. This enables lateral accelerations to be detected as yaw corrective action occurs.

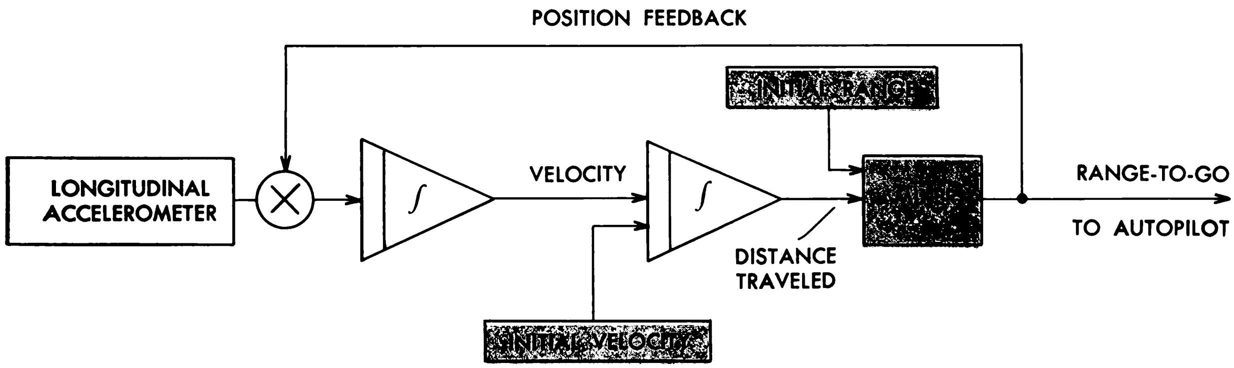
As the desired course is reached, the missile may still have a sideward velocity to the left. This could cause overshooting and oscillation about the desired course. The usual remedy for this oscillation is to dampen it by means of a rate signal as it is done in the pitch, roll, and yaw channels of a conventional attitude control system. That is, a rate signal is added to the distance error signal. However, in our case of an inertial guidance system, the rate signal is proportional to the rate of change of distance deviation, not angular deviation. The rate of change of distance deviation is the same as velocity; therefore, the output from the first integrator in the lateral channel may be fed directly to the yaw channel to reduce slow sideward oscillations.

As indicated at **B** in the illustration, it is desirable to maintain the missile on course regardless of any strong continuous existing crosswind. This requires a crab angle to be maintained by slight, but continuous, rudder deflection. In this case a special integrator control loop may be included in the guidance or control system to *support* the lateral error signal. That is, a signal based on the average lateral error and time may be provided so as to cause the integrator output to increase and command yaw rudder action until it is just sufficient to overcome drift.

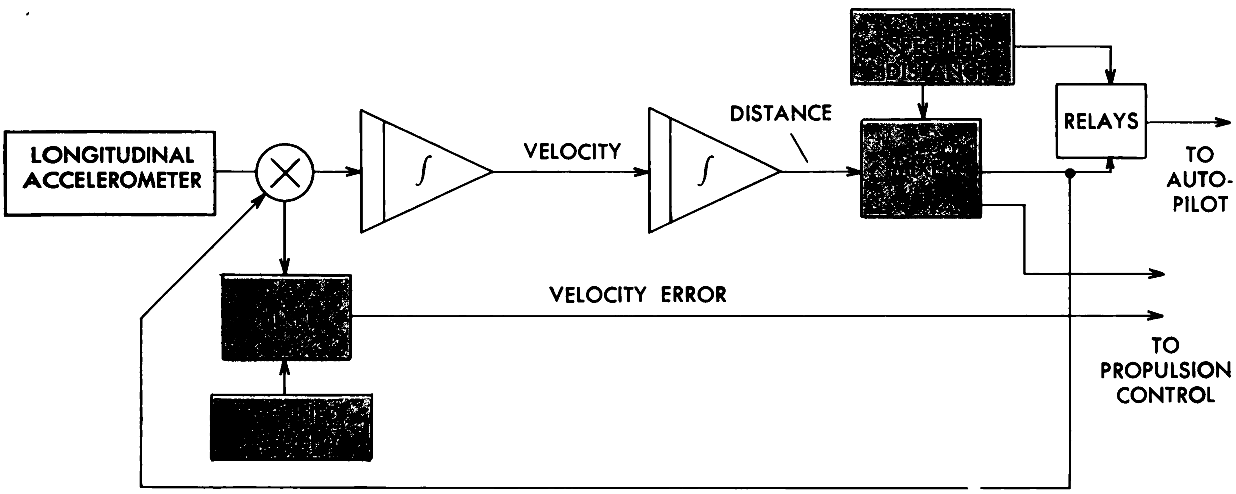
LONGITUDINAL CHANNEL. The longitudinal or range channel is similar to the lateral channel in operation. As depicted at **A** in the illustration on page 5-45, the first integrator produces an output whose magnitude indicates velocity of the craft under control. The second integrator produces an output voltage proportional to the longitudinal distance of travel. This distance must be measured from a definite starting point. That is, the distance the craft has traveled must be compared to the known distance from launch point to target. Hence, a voltage equal to the entire programmed distance the craft is to travel is set up as an initial condition at the time of launching. This voltage is combined, with opposite sign, and fed through a mixer to the output of the second integrator. In this manner, the range voltage output from the second integrator is subtracted from the initial range information, giving the *range-to-go*. Hence, the output of the longitudinal channel—the range-to-go—approaches zero volts as the flight progresses, and at zero volts the target point has been reached.



Lateral Position Error Corrected by Yaw Channel Controls



(A.) BASIC LONGITUDINAL CHANNEL



(B.) RANGE AND TIME SPECIFIED LONGITUDINAL CHANNEL

If the system does not begin operation until the craft has attained a certain velocity, this velocity must be accounted for in the computation of distance. A separate signal representing initial velocity is then fed, as shown, to the input of the second integrator. The output of the first integrator combines with the initial velocity signal to indicate the desired changes in velocity.

In some longitudinal channels it does not matter what the forward velocity of the craft is, since a velocity change simply means that the range-to-go output will reach zero either early or late. Regardless, the distance covered will be the same. On the other hand, in some inertial systems, especially those which are stellar supervised, the range-to-go must be a certain amount for every instant throughout the flight. In such a time-specified flight, output from the longitudinal channel will continually control the thrust of the propulsion unit. If the craft is ahead of schedule, it has a forward distance error of a certain amount. This will be the output of the longitudinal channel, and it will be fed to the propulsion circuits. A channel of this type is illustrated at **B** in the previous figure.

In order for forward or aft errors to be detected, desired distance must be programmed and the computed distance compared to it. This comparison may be

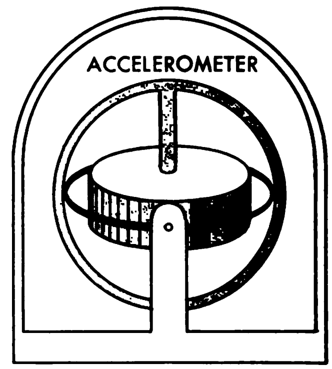
accomplished either mechanically or electronically. The difference between the signals will represent the mixer output.

Accelerometer Requirements

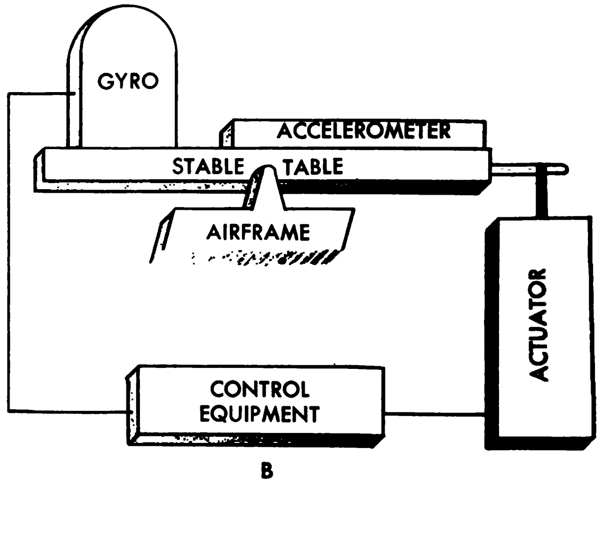
From the above discussion you can see that distance or range is the quantity which may be fed into an inertial guidance system. We arrive at this quantity by measuring accelerations. It is the only accurate measurement available without using external conditions such as the earth's terrain, stars, or radio waves. For example, any attempt to measure velocity or distance directly usually results in the use of ram air pressure which cannot give an accurate ground speed, or the use of radar or radio signals which can be jammed. Actually an accelerometer will detect acceleration with respect to *space*. Since the motions and accelerations of the earth with respect to this space are known, the accelerometer, together with a computer, can determine ground distances. More will be said about this space relationship when considering accelerometer errors.

ACCELEROMETER ERRORS DUE TO GRAVITY. The force of gravity exerts an influence on an accelerometer. The force which *pulls down* on an accelerometer cannot be distinguished from a force exerted upon the mass due to accelerating motion. We find that the most likely method of eliminating this error is to keep the accelerometer detection axis normal (90°) to the direction of gravity. This means that longitudinal and lateral accelerometers have to be in a level plane at all times. This cannot be done unless they are mounted on a platform which is entirely free from the attitude deviations of the craft to be controlled. Here again we use our stable table. This platform must be stabilized to a much finer degree than the stable table used in an attitude control system—for even very small deviations may introduce prohibitive errors. The stable table is a fundamental part of every inertial system.

ACCELEROMETER STABILIZATION. Two basic methods of providing stabilization for inertial guidance components are depicted at A and B. One method, as shown at A, is to mount the device which must be stabilized (the accelerometer in this case) directly on the gimbals of the gyro as indicated. However, in such an installation we have a large mass which the gimbals must support and control purely by the rigidity of the gyro rotor. This is a disadvantage as such a mass may easily introduce gyro errors. Also the accelerometer would not be accessible as it must be mounted within the gyro case. Furthermore, it can be stabilized about only two axes, the gyro spin axis being the exception.



A



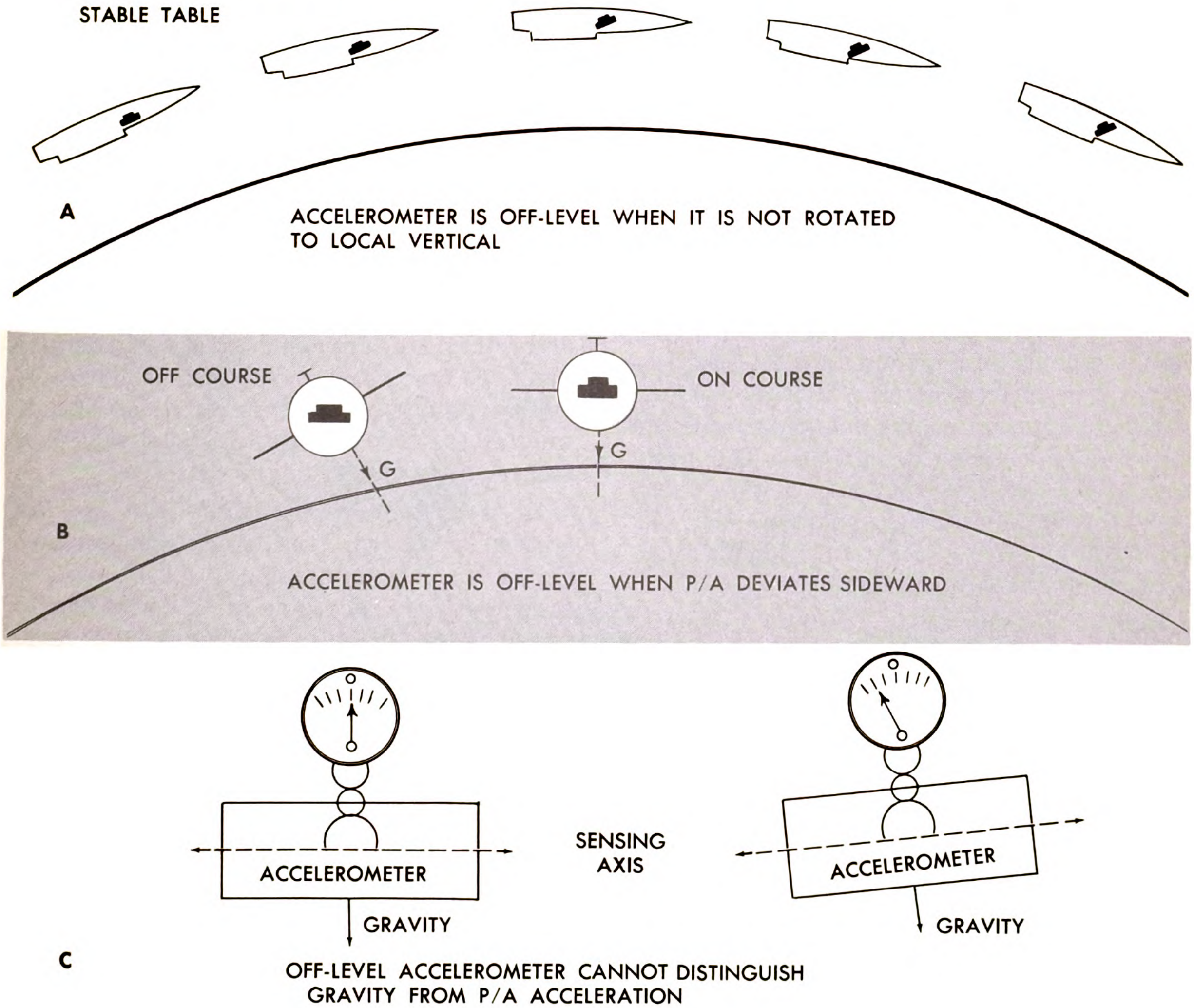
B

Methods of Maintaining a Level Accelerometer

The second and more common method used to stabilize the accelerometer is to feed a gyro signal to control equipment similar to that of an attitude stabilization channel. Such a system is illustrated in the figure at **B** and uses an actuator to move the platform about one axis of movement. Two such channels would provide complete stability. When the craft deviates in attitude the system must be fast enough to change the table-airframe relationship in order to keep the table level. The term level means perpendicular to the direction of local gravity; that is, gravity at the immediate location of the craft above the earth's surface.

If a stable table uses a vertical gyro which is not slaved to local gravity, the gyro will tend to maintain its original position during flight. This will create a movement of the table from local gravity as depicted at **A** in the illustration of accelerometer errors.

In the example illustrated, the deviation of the stable table is caused by the fact that the craft follows the curvature of the earth while the gyro tends to maintain its initial position in space. If you, as an observer were



Accelerometer Errors

on the craft, you would find that the gyro would appear to precess. In addition, even if the craft were not moving around the earth, the gyro would appear to precess due to the earth's rotation. You will remember, apparent precession was discussed in an earlier chapter. Besides that, due to the earth's rotation, we now have a second factor involving apparent precession; that is, the precession of the craft itself seen in the gyro as it proceeds around the earth. Apparent precession would produce an off-level table as it aligns with the gyro. The table will also be off-level due to sideward positional error as shown at B in the previous illustration. When the stable table is not level, the accelerometer will produce a steady signal equal to the component of gravity which affects the detecting axis, as shown at C. That is, the more it deviates from a level condition, the greater will be the signal due to gravity. This gravity signal can be cancelled by combining the acceleration output with a certain proportion of the position error. For small position errors the gravity signal will be proportional to the position error. The accelerometer will then produce a signal representing true acceleration with the gravity component removed. The position error signal is fed back to the mixer from the output of the computer—as was depicted in the lateral and longitudinal channels previously illustrated.

Inertial systems which maintain a certain average ground speed can be made to overcome apparent precession. For example, the measured velocity information in the range channel is used to precess the vertical gyro at a rate just sufficient to overcome this error.

If the gyro is slaved to vertical by means of some pendulous sensor, the problem of maintaining a level accelerometer is remedied. Unfortunately, a pendulum is not only sensitive to local gravity, but also to the craft's accelerations.

In the case of an accelerometer aligned to detect vertical accelerations, we may find that the force of gravity could not be compensated for by using a stable table. Instead, the accelerometer output may then be combined with a steady signal which counteracts the output due to gravity. Another method could be to impress an equal and opposite force on the accelerometer by some means such as a magnetic field.

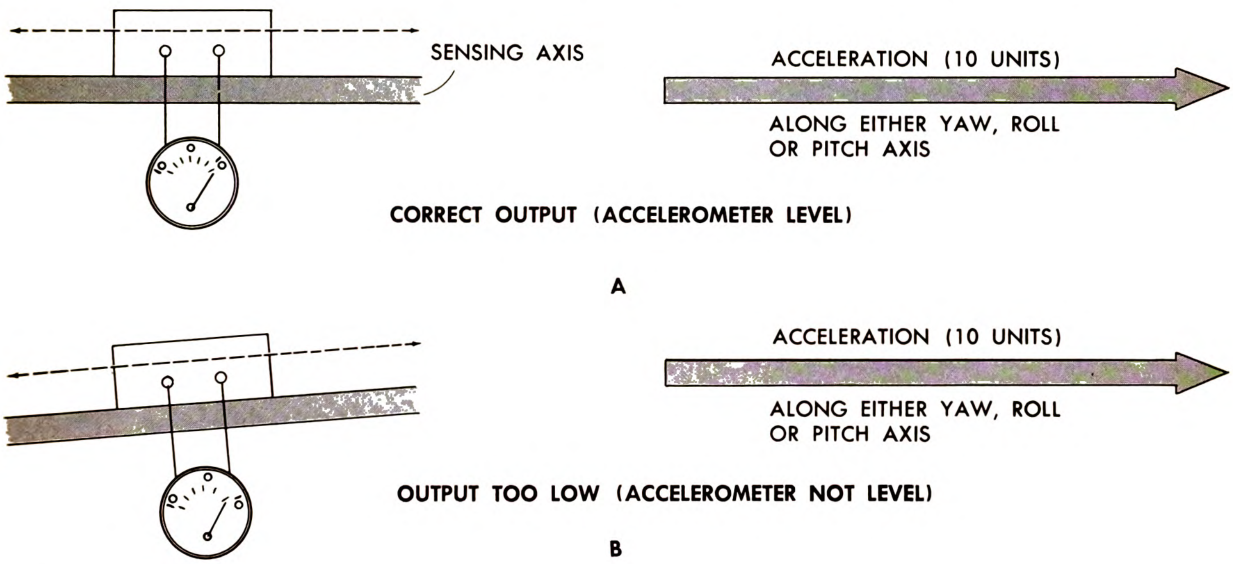
ACCELEROMETER ERRORS DUE TO MISALIGNMENT. The sensing axis of each accelerometer must be *perfectly aligned with the desired axis* in which acceleration is to be measured. If it is not, there will not only be a constant error due to gravity, but the output will not be accurate for any given craft's acceleration—as is depicted in the accompanying series of figures. When a craft accelerates or decelerates, only the component of acceleration which acts along the sensing axis will affect the output. Note at A the output shows all the acceleration since it is acting parallel to the sensing axis. At B the inertial force *along the sensing axis* is not as great. As in the consideration of gravity error, the accelerometer error will depend upon the amount of misalignment. The possibility of this error gives an added important reason for the use of stable tables to insure accelerometer alignment.

In a completely *gyro-stabilized inertial system* as shown, the accelerometers are placed on a gyro-stabilized platform. The gyroscopes are placed so as to detect errors in the pitch, roll, and yaw axes. Any off-level condition is apparent in the output of the gyros. This output is then amplified and applied to a servo mechanism in the control channel which corrects the attitude of the platform.

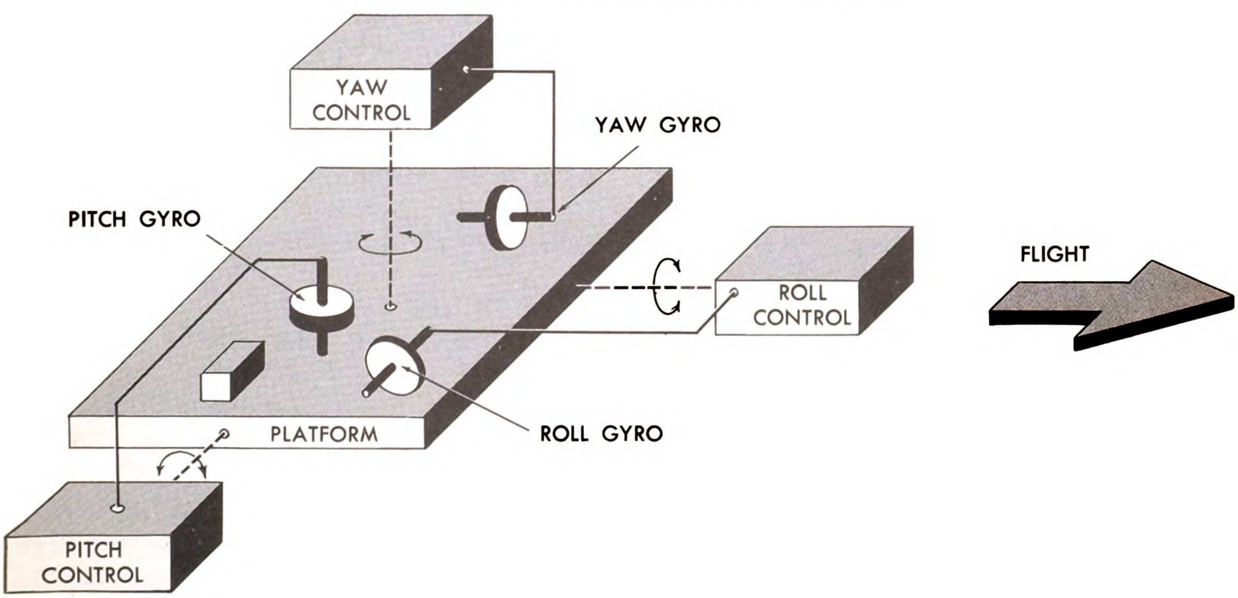
The principle of inertial guidance is simple, as is evidenced by the lateral and longitudinal channel block diagrams previously illustrated. However, the fulfillment of these principles is difficult. Compensations in addition to those which have already been mentioned are necessary, and the longer the range, the more important these errors become. The science behind these corrections is

already represented by lengthy and numerous equations of astronomy, meteorology, navigation, and physics.

Most of these errors in acceleration detection arise from the fact that an accelerometer detects *all* acceleration within a certain plane of the accelerometer. As previously mentioned, this acceleration is not with respect to the earth, but to *inertial space*. Inertial space is the condition in which all accelerations are assumed to be zero and all gravitational forces (mass attractions) of the universe are balanced. This simply means that no acceleration or mass attraction



Acceleration Errors Due to Accelerometer Misalignment



Basic Gyro-Stabilized Inertial System

exists. For no accelerations to exist, there must be no rotational motion, as rotational motion possesses acceleration on a body acting toward the axis of rotation. Rotational motion produces a change of direction of a rotating body which results in a circular rather than a linear path. The accelerometer will, therefore, detect acceleration as it rotates with the earth about the polar axis.

An accelerometer must be sensitive enough to detect minute accelerations and rugged enough to withstand extreme accelerations at launching. It should also be pressure and temperature compensated.

An accelerometer must be unidirectional. That is, it must detect acceleration (or deceleration) in just *one* direction. For example, if longitudinal accelerations affected the output of a lateral accelerometer even to a small extent, it would be useless. Accelerometers must also remain unaffected by rotational accelerations.

An accelerometer must respond uniformly to all accelerations with a frequency from zero to about 35 cycles per second. This is the same as stating that for a certain amount of acceleration, the output must be the same regardless of how fast that acceleration may be changing. It is desirable for the accelerometer weight and size to be a minimum. Not only is this true to help keep the controlled craft small in size, but it also simplifies accelerometer orientation.

GYRO ERROR. The gyros intended for guidance must be of much greater accuracy than those used for stabilization. For example, assume that the directional gyro of an *attitude control* system has an inherent error of 1/2 degree. It would stabilize a craft in yaw with a slightly incorrect heading.

However, as soon as this heading error was revealed as a lateral distance off-course error, the inertial guidance signal to the yaw channel would override the control system to head the craft back on course. (If a stable table is used as the attitude reference, there would be no additional directional gyro needed in the control system.)

Suppose we compare this to a *guidance* gyro which had within itself an error of one minute of arc (one *sixtieth* of a degree). A target error would be produced. Or in other words, in a flight of several hundred nautical miles this error of one minute of arc may result in a lateral *distance* error of over one-eighth nautical mile from the target.

INTEGRATOR ERROR. Any error in forward acceleration is accumulative as time progresses. Consider the average acceleration required to displace an object a distance s in t seconds. Assuming zero initial velocity, we have:

$$s = 1/2\ at^2$$

where s is the distance in feet;

a, the acceleration in ft/sec/sec; and

t, the time in seconds.

The accelerometer sensitivity required for a given system accuracy may be found by the use of this equation. For example, if a craft has a velocity of 1,500 mph, the required time to travel 400 miles is 960 seconds. An acceleration error at the beginning of flight, which would cause an error of 3,000 feet at the target, may be found by substituting the time and distance into the formula. This would give the acceleration error to be 0.0065 ft/sec/sec, assuming all other errors are zero. Transposing the formula given previously, $s = 1/2\ at^2$, we have:

$$1/2\ a = \frac{s}{t^2}$$

$$a = \frac{2s}{t^2}$$

$$\frac{2 \times 3000}{921{,}600} = 0.0065\ ft/sec/sec$$

If the flight time is increased, the target error will increase as the square of this time for any given acceleration error at the beginning of flight. This is evidenced by the previous formula in which distance depends on the square of time. If a represents a cer-

tain *amount of error* due to the acceleration, then s represents the *distance error* due to the acceleration error.

The fact that a distance error builds up with the square of time is also affirmed by considering operation of the two integrators in each channel. You will find that an error in the first integrator will produce a velocity output with an error which steadily increases. This steadily increasing error is then fed to the second integrator. Here the error in the output increases at a faster and faster rate. A graph of this error would be a curved line which becomes steeper and steeper.

Consideration of Initial Conditions

Throughout this discussion, reference has been made to distance, velocity, and accelerations with respect to a starting point. The starting point is the location at which the inertial system begins to *keep track* of accelerations and begins to compute distances and velocities. This starting point may or may not be the launch point.

Many *initial conditions* must be known in order for the inertial system to guide a missile to a target. In the lateral channel the exact distance to right or left of the desired course, the drift (sideward velocity), and the sideward acceleration must be sensed and fed into the system. In the longitudinal channel the same three factors, initial position, velocity, and acceleration must be detected. The initial position refers to the total distance along the course to the target. If the system is of the type which uses a programmed course, then the initial position will be the distance between the actual position and the desired position. If an inertial channel is used to control movements along the pitch axis, the same three initial conditions are required for that channel.

The other references are of vital importance. One, an exact azimuth reference must be established. The effect of any inaccuracy in this reference has already been discussed. Two, the exact direction of local gravity must be established. The importance of this was confirmed in the discussion on acceleration errors due to gravity and accelerometer misalignment.

GROUND LAUNCHING. Suppose we consider first a ground launched missile. If the starting point is at the launcher, then all of the initial conditions at this starting point *will be zero*. The accelerometers must accurately detect every bit of launching acceleration and the direction in which it occurs in order to furnish information for the completion of the flight. An alternative is to establish a starting point sometime after the launching phase is completed. Initial position, velocity, and acceleration data could then be fed into the channels and relayed to the missile by ground tracking radar, computers, or radio relay link.

AIR-TO-AIR LAUNCHING. If a missile is air launched from an aircraft, the problem of determining initial conditions and relaying them to the guidance system is greater. In this case it is harder to determine the accurate position of the launcher with respect to the target when the launcher is rapidly moving. The velocity and the acceleration of the launching aircraft with respect to the ground must be accurately determined. The equipment for doing this must be located in the launching aircraft without the benefit of the natural stable reference of the ground.

There are several methods of indicating a certain quantity such as distance, velocity, or acceleration within a missile. Often the position of a potentiometer contact (or selsyn rotor) is used. The resulting voltage will represent the amount of velocity at, or some distance from, the starting point. Distance can also be easily indicated by the angular position of a gear drum or other rotating mechanism.

Before a missile is launched from an aircraft, the potentiometer or gear drum may be positioned by an electrical drive mechanism wired to the aircraft. If the missile is the type in which the starting point occurs after it is on its own, the potentiometer or

gear drum must be positioned by radio signals. Normally the radio signal will be in the form of pulses which frequency- or amplitude-modulate a carrier. These pulses are demodulated in the missile receiver. Each pulse is made to step the potentiometer or gear drum a small amount. The number of transmitted pulses may determine the position of the indicating device.

INERTIAL COMPUTER INITIAL CONDITIONS. Once initial conditions have been indicated, they must influence the computer. For example, it would be impossible for the computer to continually produce a solution of the missile ground speed if the velocity at starting point were not known. Similarly, the computer could not produce a continual solution of distance covered if the initial distance were not accounted for.

The method of inserting this information will depend mostly on whether the inertial system is of the programmed type. The programmed inertial system, you will remember, specified either velocity or distance for every period of time during the flight.

If the inertial system is not programmed, the distance signal will represent range-to-go and will be either a large voltage or a certain gear or drum position. It will exist with the position information at the output of the second integrator. The velocity signal will be proportional to the total velocity of the missile or to the deviation from a preset velocity and will be combined with the computed velocity.

In a programmed system the initial conditions represent deviations from the desired velocity or position. Initial signals may be inserted in such a manner that the signals *cancel out* as the missile acquires the specified trajectory.

For example, suppose the missile were a mile behind the desired position at the starting instant. A voltage representing this mile must be applied to the longitudinal channel. The first thought would be to combine it with the output of the second integrator which represents distance ahead of or behind the desired position. However, if you refer to the previous longitudinal block diagrams, you will find that this output goes to the propulsion circuits to vary the speed of the missile. The signal representing this one mile error would increase the speed until the distance error no longer exists. However, when this happens the missile may have reached a velocity above normal and begin to go beyond the specified position. This may, however, produce an output of the second integrator which slows the missile down. Such a condition would produce oscillation forward and aft of the specified position since either the velocity or the position is not correct. In the lateral channel a sideward oscillation would result in the same manner if an initial distance error existed.

To avoid this oscillation, the distance signal is applied at an earlier point in the computer, such as the input to the first or the second integrator. This will produce an error signal which will gradually cancel out due to the position feedback paths which exist in the programmed type of inertial channels.

Oscillation could also result if initial velocity is combined directly with computed velocity. Suppose an initial velocity error signal of excessive speed is applied to the input of the first integrator. This signal would be applied to the input of the second integrator to eventually produce a forward distance error signal. This distance error signal would slow the missile down. However, by the time the missile slows down enough to null the distance error, it is already too slow. This will produce an aft distance error which again accelerates the missile and the process repeats itself, producing oscillation. Therefore, a velocity signal must be applied to the missile in the manner similar to that for a distance signal. Initial velocity and distance deviations will, in most cases, be inserted into the inertial channels in the form of voltages which are combined at the proper point and in the proper ratio and sense with the integrator outputs and inputs.

CHAPTER 6

RADAR FUNDAMENTALS

One of the first observations of "radio echoes" was made in the United States in 1922 by Dr. Albert H. Taylor of the Naval Research Laboratory. Dr. Taylor observed that a ship passing between a radio transmitter and receiver reflected some of the waves back toward the transmitter. Between 1922 and 1930, further tests proved the military value of this principle for the detection of surface vessels which were hidden by smoke, fog, or darkness. Further developments were conducted with carefully guarded secrecy. During this same period Dr. Breit and Dr. Tuve of the Carnegie Institute published reports on the reflection of pulse transmission from electrified layers in the upper atmosphere which forms the earth's ceiling. This led to the application of the principle to the detection of aircraft. Other countries carried on further experiments independently and with the utmost secrecy. During 1936 the United States Army was engaged in the development of a radar warning system for coastal frontiers. The pulse system of transmission was further developed between 1936 and 1940, and mass production of radar equipment was under way.

Radar first made history in that phase of World War II known as the Battle for Britain. Outnumbered more than ten to one, the Fighter Command of the Royal Air Force was able to deal the German Air Force its first smashing defeat. This was accomplished by locating the enemy bombers with radar and having fighter aircraft waiting at the right place to intercept them, no matter when or from what direction they attacked.

RADAR SYSTEMS

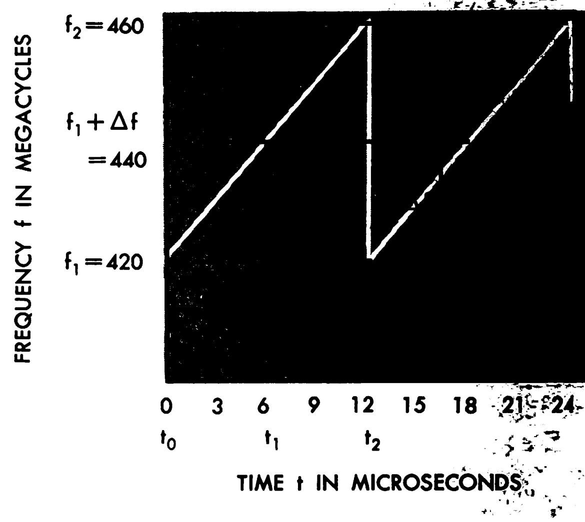
A radar system consists of a transmitter which sends out radio signals, a receiver which is located at the same site, and an indicator which gives a visual indication of echoes returned from the target.

When a radio signal which has a constant frequency is emitted by a radio transmitter, radio waves travel out in all directions in a manner similar to light and sound waves. They are reflected by any object they may strike. On striking the object, components of the wave are reflected and travel in all directions. Some of the reflected waves return to the site of the transmitter originating them, where they are picked up by the receiver, providing it is tuned to the correct frequency. This is a simple explanation. Naturally, there are some problems which must be solved. First, there is the problem of having a powerful transmitter near the receiver and operating on the same frequency. In order for the receiver to detect the reflected signal and indicate the presence

of the target, the transmitter signal must be prevented from affecting the receiver.

The second problem is measuring distance. If a continuous constant frequency signal was transmitted, it would be impossible for the receiver to distinguish between various echoes at different distances, for they would all be alike. Some means must be provided to eliminate the problems involved in using constant frequency signals.

There are several systems which satisfactorily solve these problems. Two of these systems are the frequency-modulation system and the pulse-modulation system. The pulse-modulation system is the most important of the two.



Frequency-Modulation Chart

Frequency-Modulation System

Because each cycle of a frequency-modulated wave differs by a small increase in frequency from the others of that wave, a frequency-modulation system makes it possible to identify each cycle of a wave transmitted and to recognize it from all others when it returns to the receiver. With the designing of a transmitter which produces a signal that regularly changes over a known range of frequencies, it is possible to identify any particular reflected signal cycle.

An example of a frequency-modulation signal is plotted against time in the frequency-modulation chart. As shown, the 420 mc frequency increases linearly to 460 mc and then quickly drops to 420 mc again. When the frequency drops to 420 mc, the frequency cycle starts over again.

Since the frequency regularly changes 40 mc with respect to time, you can use its value at any time during its cycle as the basis for computing the time elapsed after the start of the frequency cycle. For example, the transmitter sends a 420 mc toward an object, strikes the object and returns at time t_1 when the transmitter is sending out a new frequency of 440 mc. Since f changes from 420 mc to 460 mc in 12.4 microseconds, the time required for the 420 mc signal to change from 420 to 440 mc is 20/40 of 12.4 or 6.2 microseconds. Radar waves strike and return from an object one nautical mile away in 12.4 microseconds. Therefore the distance of the object in the example is 6.2/12.4 or 1/2 nautical mile.

In the frequency-modulation system, two separate signals are fed to the receiver at the same time. For example, at t_1, the 440 mc transmitted signal and the 420 mc reflected signal reach the receiver simultaneously. When these two signals are mixed in the receiver a beat note results. The frequency of the beat note varies directly with the distance to the object, increasing as the distance increases. A device that measures frequency can be calibrated to indicate range (distance to object).

Pulse System

The pulse system of detection is employed in almost all radar sets. In this system, the transmitter is turned on for short periods and off for long periods. During the periods when the transmitter is turned on, it transmits a short burst of energy called a pulse. When a pulse strikes any object, part of the reflected energy is returned to the receiver, where it is displayed on the screen of a cathode ray tube. (The cathode ray tube is a device capable of measuring periods of

time as short as one-millionth of a second.) Since the transmitter is turned off after each pulse, it does not interfere with the receiver (as would be the case if a constant signal were used).

Complete location of an object in space by the radar pulses depends upon two factors, the range or distance of the target, and the direction, including both the azimuth and elevation directions of the target.

DETERMINING RANGE. The successful use of a pulse-modulated radar set depends primarily on its ability to measure distance in terms of time. When radio energy is radiated into space, it continues to travel with a constant velocity. Its velocity is that of light, or about 186,000 statute miles per second or 162,000 nautical miles per second. In more useful terms, radio waves travel a nautical mile in 6.2 microseconds. They travel a radar mile, that is, go-and-return mile, in 12.4 microseconds.

This constant velocity is used in radar to determine the distance or range of a target by measuring the time required for a pulse to travel to a target and return. Suppose, for example, a pulse of radio energy is transmitted toward a target some distance away and the radar echo returns after 620 microseconds. Energy moves a mile and back or a radar mile in 12.4 microseconds. The distance of the target is 620/12.4 or 50 nautical miles.

In order to use the time-range relationship, a time-measuring device must be used. The cathode ray tube is useful for this purpose since it responds to changes as rapid as one microsecond apart. A time base is provided by using a linear sweep to produce a known rate of motion of an electron beam across the screen of the cathode ray tube.

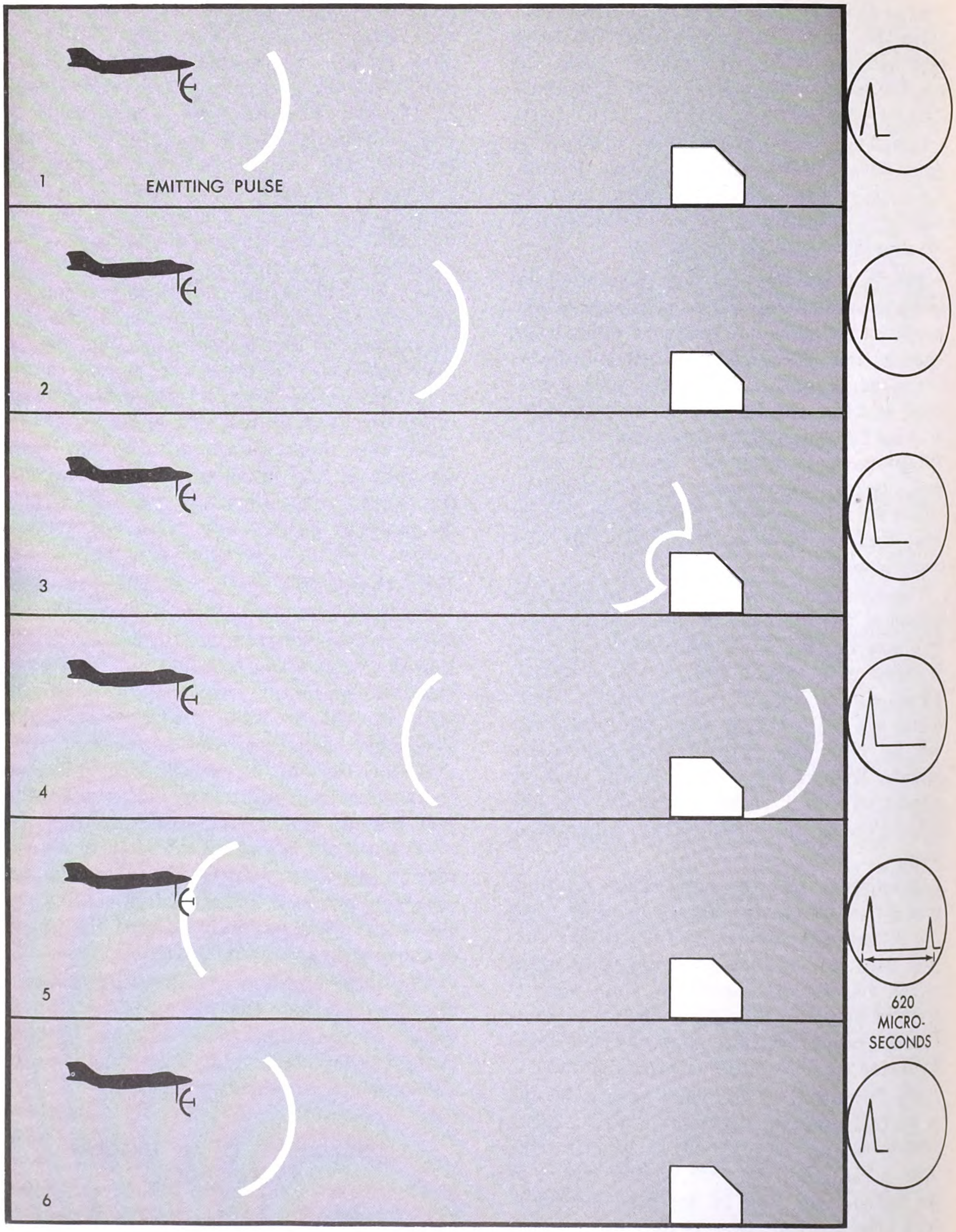
The formation of the time base is shown in the following illustration. In (1) a radar pulse is leaving the airplane. At the time the pulse is radiated, the spot on the screen of the cathode ray tube is deflected vertically for a brief instant, then it continues across the screen to the right. In (2) and (3) the pulse is traveling toward the target, and the spot is moving across the screen. When the pulse strikes the target, there is no deflection, since the energy is at the target itself. In (4) the reflected pulse is returning. In (5) the reflected energy has returned to the receiver, and there is a second vertical deflection of the spot on the right side of the cathode ray screen. The distance between the two upward deflections serves as the basis for determining the range of the target from the radar antenna. Let us assume, for example, that the set is designed so the spot moves across the cathode ray tube in 700 microseconds. The spot was almost to the end before the echo arrived, its position indicating that it took 620 microseconds to reach that point. Since radar waves travel one mile in 12.4 microseconds, the range of the target in the illustration is 620/12.4 or 50 nautical miles away. The last picture shows another pulse being emitted and the start of the formation of a new time base.

POSITION DETERMINATION. Two dimensions must be considered in determining the position of a target. One is azimuth, which is the relative horizontal direction of the target with respect to some direction reference expressed in degrees. As an example, this direction may be expressed with reference to true north if the radar set is a ground installation, or with reference to the aircraft if the set is airborne. The other dimension is elevation which, like azimuth, can also be expressed in degrees. Elevation expresses the angular degrees that the target is above or below the radar set.

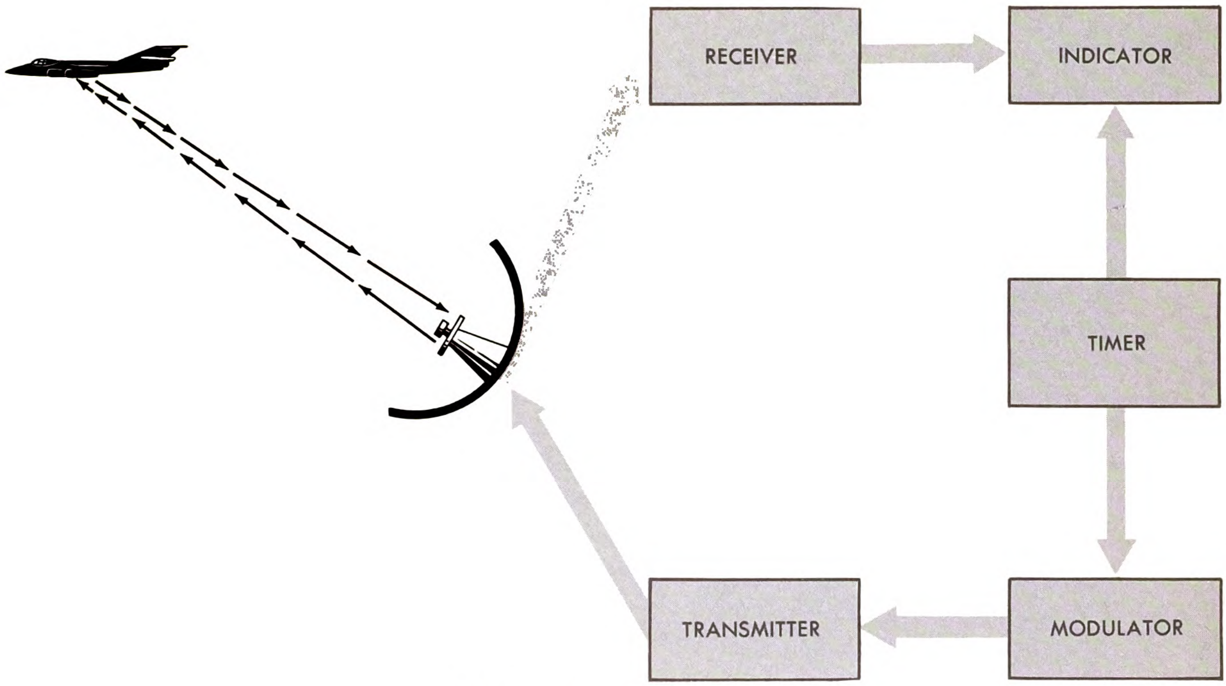
The determination of azimuth and elevation depends upon the directional characteristics of antennas and antenna arrays. Antennas for performing these functions are discussed later in this chapter.

COMPONENTS OF AN AIRBORNE RADAR SYSTEM

The six basic components in a typical airborne radar system are the timer, modulator,



Formation of a Time Base



A Basic Airborne Radar System

transmitter, antenna, receiver, and indicator as shown in the illustration.

The Timer

The timer, or synchronizer, is the heart of all pulse radar systems. Its function is to insure that all circuits connected with the radar system operate in a definite time relationship with each other, and that the interval between pulses is of the proper length. The timer may be a separate unit by itself, or it may be included in the transmitter.

The Modulator

The modulator is usually a source of power for the transmitter. It is controlled by the pulse from the timer. It is sometimes called the keyer.

The Transmitter

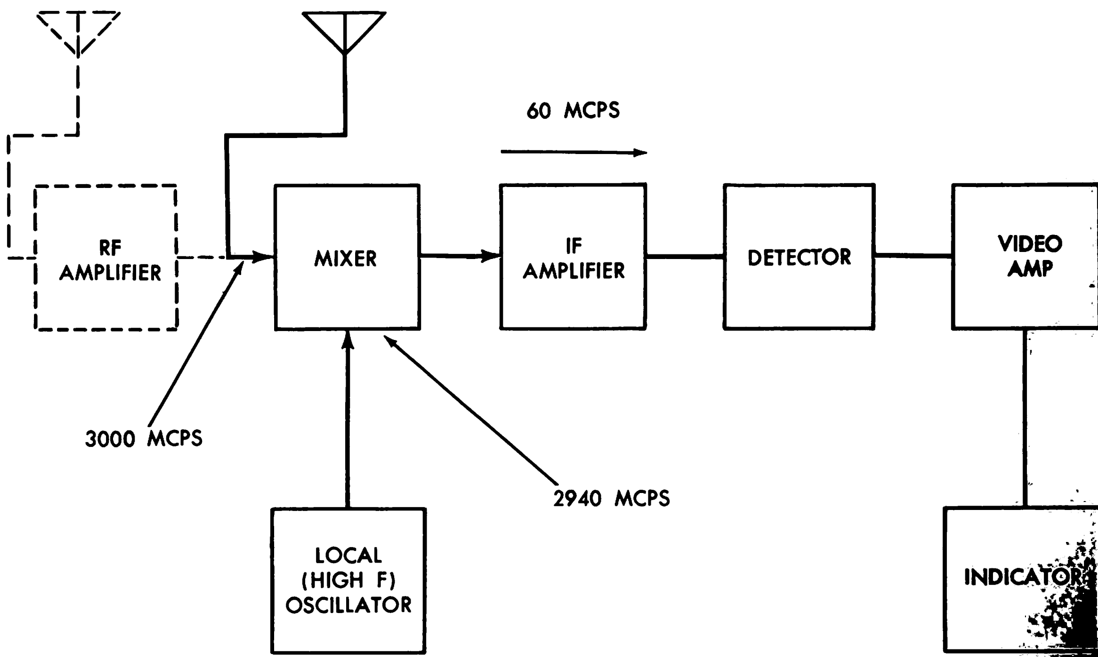
The transmitter provides RF energy at an extremely high power in very short pulses. The carrier frequency must be extremely high to get many cycles into a short pulse.

The Antenna

The antenna is very directional in nature because it must measure the angles of elevation and the bearing of the target. To obtain this directivity at centimeter wavelengths, ordinary dipole antennas with parabolic reflectors are used. Usually, the same antenna is used for both transmitting and receiving. When one antenna is used, some kind of a switching device is required for connecting the antenna to the transmitter when a pulse is being radiated and to the receiver during the interval between pulses. Since the antenna only sees in one direction, it is usually rotated or moved about to cover the area in front of the radar set. This is called searching. The presence of targets in the area is established by searching.

The Receiver

The receiver in radar equipment is usually a superheterodyne receiver. It must be capable of accepting signals in a bandwidth of



Block Diagram of a Superheterodyne Receiver

one to ten megacycles. The receiver takes the weak RF echoes from the antenna, amplifies them, and changes them to video signals which can be applied to the indicator.

FREQUENCY CONVERSION. Many stages of amplification are required in a radar receiver in order to make a very weak signal echo voltage at the antenna give a good indication on the radar screen. In the simplest type receiver, there are several RF stages followed by a detector and by a video amplifier. In spite of this, the gain (amplification) and selectivity of an RF amplifier decrease as the operating frequency increases. This makes it desirable to convert the high radio frequency to a lower frequency, and then to amplify the low frequency. Furthermore, it is easier to design RF amplifiers with a higher gain when the frequency is constant. The modern receiver employs a high-gain and highly selective amplifier in the IF stage designed to operate at one frequency to which all received frequencies are converted.

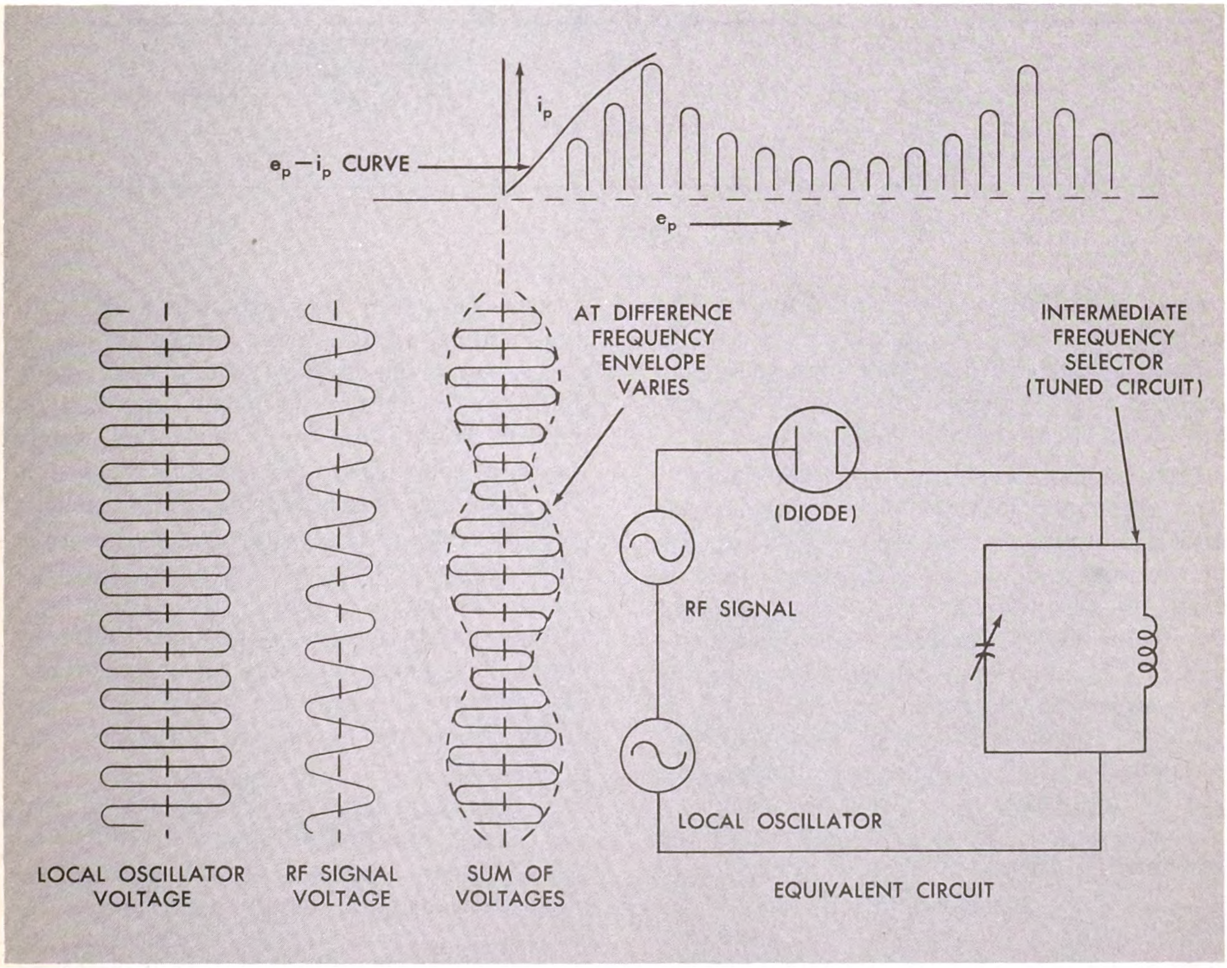
IF AMPLIFIER. Frequency conv[illegible] accomplished by mixing the RF [illegible] local oscillator frequency in a s[illegible] odyne receiver. This mixing [illegible] frequency equal to the differen[illegible] two frequencies. If the two freq[illegible] nearly the same, the difference [illegible] will be low, but not as low as the [illegible] quencies, and is called the interme[illegible] quency (IF). If the RF is [illegible] mcps, it is usually amplified [illegible] amplifier before being fed to the [illegible] higher, the RF signal is introduc[illegible] into the mixer. After mixing, [illegible] quency is amplified by an IF [illegible] then is detected in the same ma[illegible] other signal.

THE MIXER. Production of a [illegible] quency in the mixer stage requi[illegible] a nonlinear electrical circuit fo[illegible] tion of a new frequency. The [illegible] circuit is called, produces the new [illegible] in the same way as a modulator o[illegible] This stage is called the first detector[illegible]

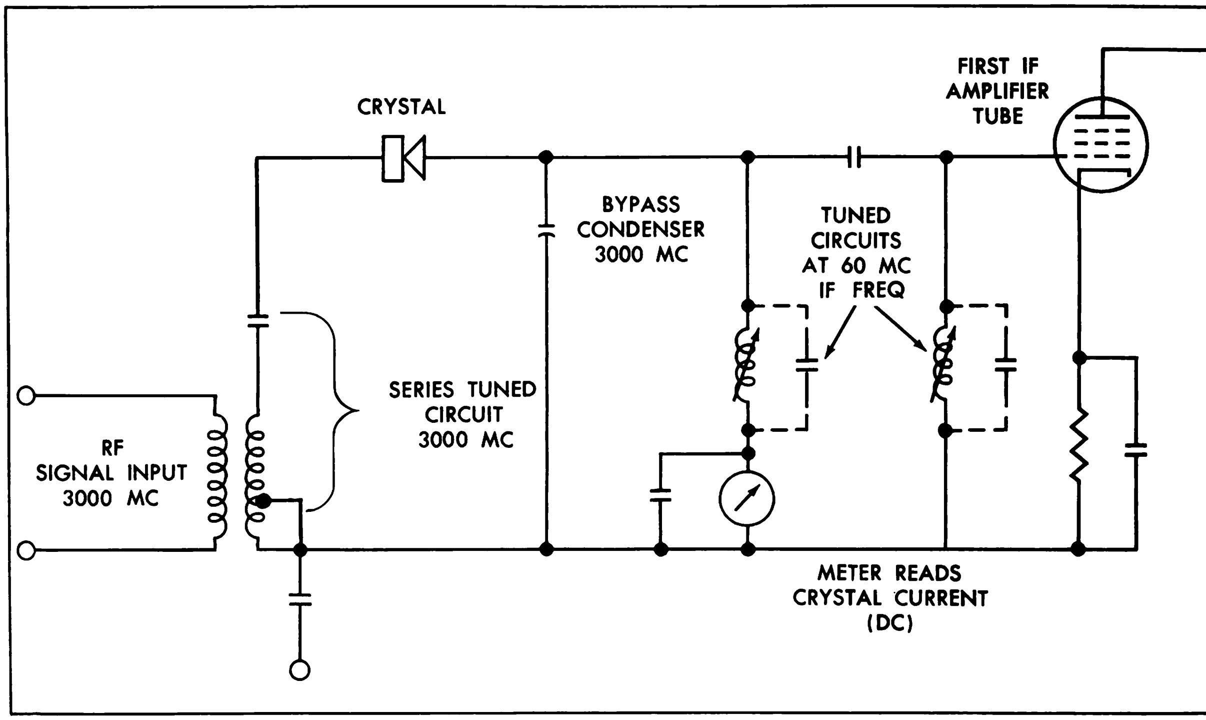
When two slightly different frequencies are added as shown in the illustration, the waveshape representing the combining of the local oscillator frequency and the RF signal contains only these two frequencies. After distortion by the mixer the waveshape contains not only these two frequencies, but a frequency equal to the sum of the two frequencies and another equal to the difference between the two frequencies. Since it is the difference frequency that is desired, the tuned circuits in the IF amplifier select this frequency and amplify it. If the RF signal is modulated, the difference frequency will also be modulated in the same manner.

In the detector circuits of low-frequency radar sets, the diode is sometimes used. In high frequency sets, however, the transit time in the tube and the thermal noise it introduces makes the vacuum tube inferior to the crystal.

Most 10-cm and all 3-cm sets use the silicon crystal for the nonlinear element in the mixer stage. In the simplified crystal mixer circuit shown, currents from both the RF signal and the local oscillator signal flow through the crystal. The difference frequency is selected by the tuned circuits, which are connected to the grid of the first amplifier tube. Any nonlinear device produces a direct



Mixer Action



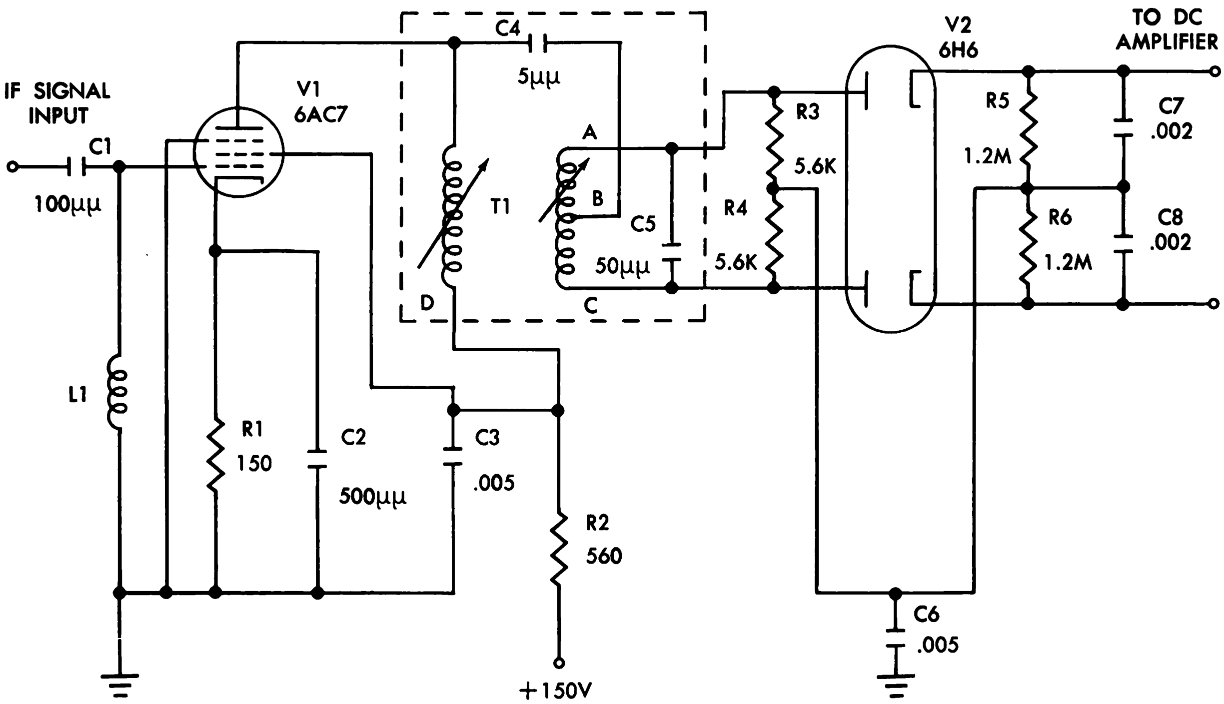
Crystal Mixer

current which may be read on a meter. The adjustment of the oscillator for maximum output is determined by observing the meter reading which varies directly with the amplitude of the oscillator signal.

AUTOMATIC FREQUENCY CONTROL (AFC). Any frequency drift of the local oscillator causes the intermediate frequency to change by the same amount. To compensate for this drift, the IF bandwidth may be increased in the design of the receiver. This increase in bandwidth will raise the noise level, but if the increase is not made, the pulse will be distorted owing to the loss of high frequency components. The performance of the receiver can be improved by automatic frequency control which avoids both difficulties. If the intermediate frequency changes for any reason, the AFC circuit brings it back to its proper value by tuning the local oscillator.

In the circuit of the frequency discriminator, the input signal is at an intermediate frequency of 60 megacycles. A drift of the intermediate frequency produces a DC voltage change which, after amplification, is impressed on the repeller grid of the klystron oscillator to retune it to the correct frequency. Tube *V1* is an IF amplifier, where the input comes from the regular IF channel. The secondary of transformer *T1* is tuned to resonance at the IF frequency. The coupling of the primary and secondary and the tuning of the primary are adjusted to give a voltage across the secondary that differs by 90° in time phase with the primary voltage. The primary is connected to the center tap of the secondary by means of the coupling condenser *C4*.

Notice the vector diagrams and equivalent circuit of a discriminator. As shown in the equivalent circuit, (1), the voltage applied between the upper diode plate and cathode is the drop across the resistor between *A* and *D*. This is the vector sum of the voltage across the upper half of the secondary *A* to *B* and the voltage primary *B* to *D*.



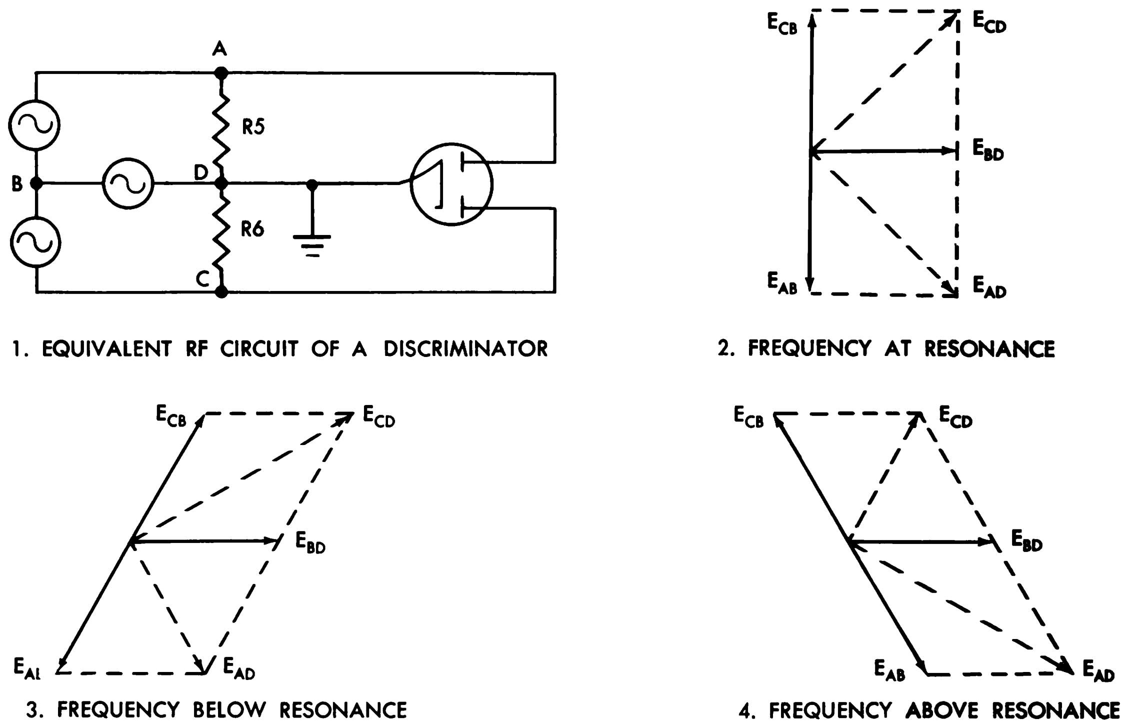
Frequency Discriminator

Similarly, the voltage applied to the lower diode plate is that across the lower half of the secondary *C* to *B* plus the primary voltage *B* to *D*.

When the IF signal is at its proper frequency, the voltage across the secondary, A to C, is 90° out-of-phase with the voltage across the primary, *B* to *D*. Thus, in illustration (2), E_{CB} leads E_{BD} by 90° while E_{AB} lags E_{BD} by 90°. Since the secondary is center-tapped to make E_{AB} equal to E_{CB}, the vector sums are equal in magnitude. Equal signals on the two diode plates produce equal currents in the cathodes, which in turn produce DC voltage drops across *R5* and *R6* which are equal but of opposite polarity. The output to the DC amplifier is therefore zero (see the frequency discriminator circuit).

If the IF signal changes in frequency, the secondary circuit is no longer tuned to resonance, and the voltage *A* to *C* no longer differs by 90°. If the frequency decreases, a lag of more than 90° is produced; if the frequency increases, a lag of less than 90° occurs. If the former case the voltage applied to the lower diode plate is greater, and the output to the DC amplifier is negative. In the latter case conditions are reversed, and a positive output is produced. After being amplified, the DC voltage is used to correct the local oscillator frequency.

AUTOMATIC GAIN CONTROL (AGC). The amplitude at the input to the receiver of the echo from a particular target may vary because of fading and changing of the position of the target at more or less rapid rate. If the receiver has a constant gain, there is a corresponding variation in the amplitude of the output. If this changing amplitude is objectionable, as in systems where the signal is used for automatic tracking, an automatic gain control may be used to give greater gain when the signal is weak than when it



Vector Diagrams and Equivalent Circuit of the Discriminator

is strong. Thus, the amplitude at the output is maintained relatively constant.

A common type of automatic gain control uses the signal voltage to control the bias on one or more tubes. A DC voltage obtained by rectifying the signal is used for this purpose. If the signal is too strong, the grid of the amplifier tube is made more negative to reduce the gain. A remote cut-off tube is used to prevent excessive distortion.

The voltage developed from the signal may be used to control the gain by letting it regulate the supply voltage for the plates and screen grids of the amplifier tubes. When this method is used, several stages of DC amplification are needed to furnish the AGC voltage, the last stage being a power amplifier.

In radar sets where automatic gain control is used, the voltage is developed from a particular target under observation, since the signals from all targets vary independently of each other. For this reason, a signal selector circuit chooses the interval of time which includes the echo from the desired target. Only the signal received during this brief time interval influences the receiver gain. Thus, echoes from targets at ranges appreciably different from that of the selected target have no effect on the gain.

AUTOMATIC NOISE LEVEL (ANL). The function of this circuit is to control the amplification of the receiver while the radar set is searching for a target. It keeps the noise generated in the tubes of the frequency converter and receiver section at a constant, predetermined level so that weak signals are not lost in the noise.

Limiting the noise to a constant, low level permits a tremendous amount of amplification to be built up into the receiver section without the danger of noise being built up to a point where the radar set will lock on it.

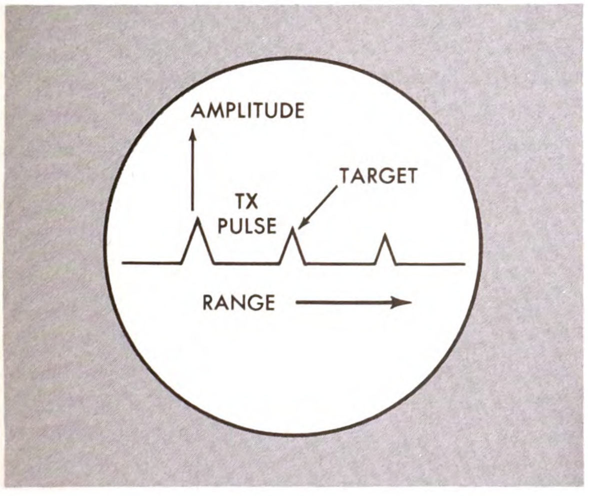
The Indicator

The indicator presents visually on the indicator screen all the necessary information to

locate the target. The method of presenting the data depends on the purpose of the radar set. Since the spot scans the indicator screen to present the data, the method of presentation is often referred to as a type of scan.

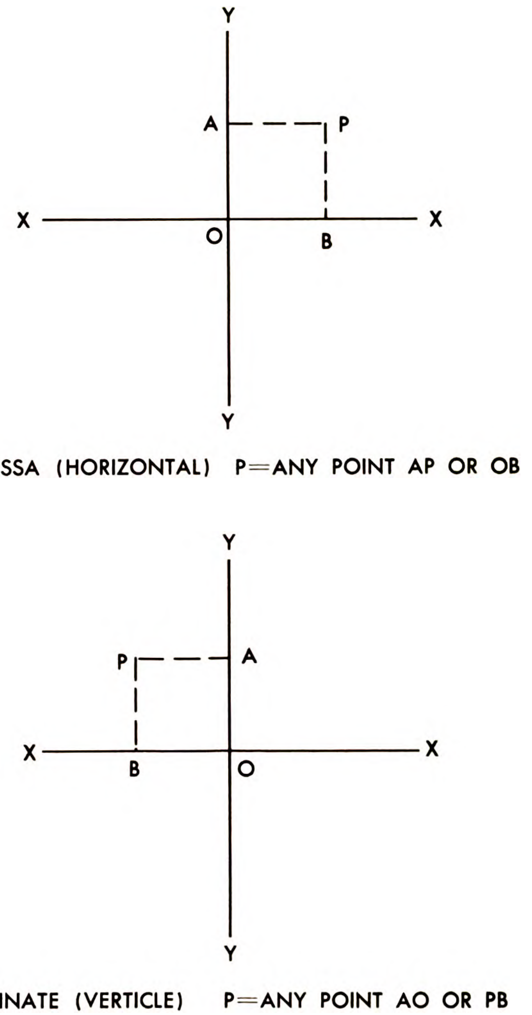
The following types of presentation used in radar receivers are illustrated.

Type A Scan. In the type A scan, the spot maintains a constant intensity, starting at the left side the instant a pulse of energy is radiated by the transmitter and traveling at a constant speed across the face of the indicator. When the spot reaches the right side of the indicator it is blanked out, flies quickly back to the left side, and repeats the process. Receiver echoes cause a vertical deflection of the spot, proportional to the strength of the received signal. The horizontal distance on the screen between transmitted pulse and echo represents the distance to the target. Although the principle function of this scan is to determine the distance of the object, it is possible to obtain a rough approximation of direction by rotating the antenna until the maximum echo is received.



Type A Scan

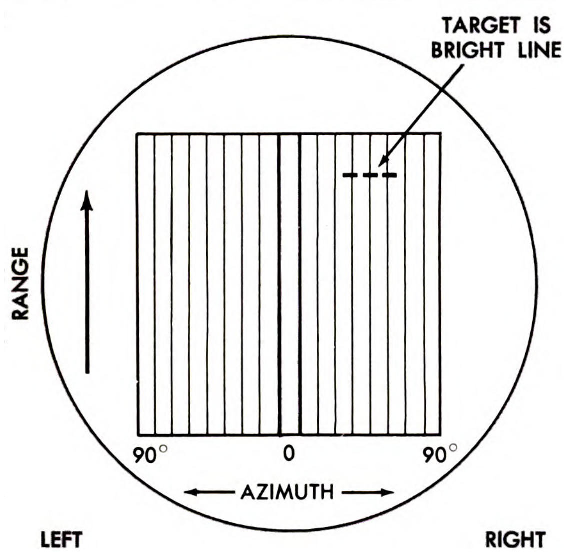
Type B Scan. In the type B scan, the bearing and range of reflecting objects are presented as abscissa and ordinate, respectively.



Abscissa and Ordinate

In this system, a highly directional antenna system is rotated about a vertical axis which causes the radiated beam to cover the horizontal plane. The spot on the screen has a horizontal motion which corresponds to at least a part of the angle of rotation of the antenna system. In the absence of other deflection the scanning spot scribes a bright horizontal line across the lower portion of the indicator. This line represents the transmitted pulses and is the so-called baseline of the pattern. A uniform vertical motion from bottom to top of the screen is also

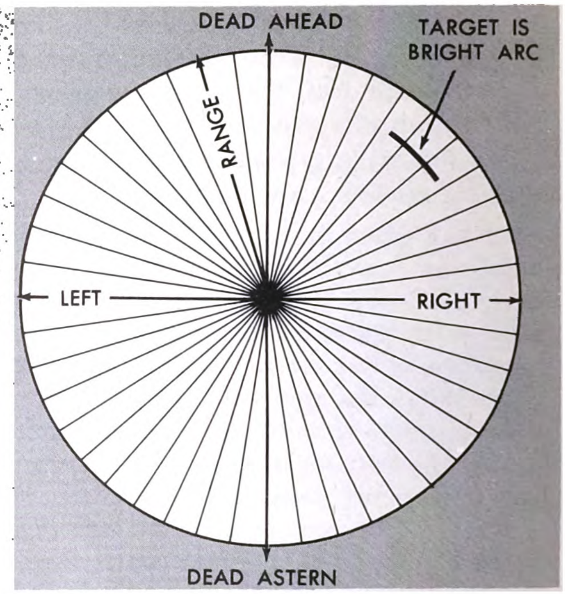
given to the scanning spot, each vertical line being synchronized with a transmitter pulse for indication of range. The vertical sweep is repeated very much more rapidly than the horizontal sweep, and the spot is maintained at low intensity. When an echo is received, the signal is impressed on the control grid of the indicator, causing a bright spot to appear on the screen. The position of this spot to the right or left of the center line of the screen indicates the azimuth of the target. The height of the spot above the baseline indicates the range of the target.



Type B Scan

THE PLAN POSITION INDICATOR (PPI) The PPI scan is another type of scan for presenting range and bearing (direction) information. You can think of the PPI scan as a modified type of B scan in which the rectangular coordinates are replaced by polar coordinates.

The antenna generally is rotated uniformly about the vertical axis so that searching is accomplished in a horizontal plane. The beam is usually narrow in azimuth and broad in elevation, and large numbers of

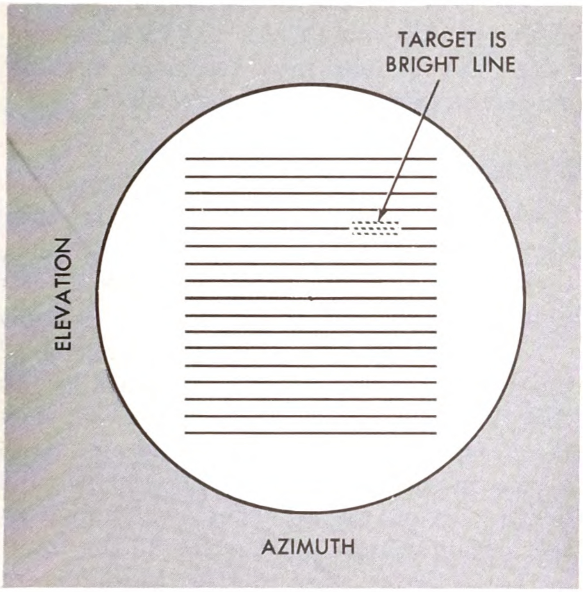


Type P (PPI)

pulses are transmitted for each rotation of the antenna. As each pulse is transmitted, the unintensified spot starts from the center of the indicator and moves toward the edge along a radial line. Upon reaching the edge of the indicator, the spot quickly jumps back to the center and begins another trace as soon as the next pulse is transmitted. As the antenna rotates, the path of the spot rotates around the center of the indicator screen so that the angle of the radial line on which the spot appears indicates the azimuth of the antenna beam, and the distance out from the center of the indicator indicates the range. When an echo is received, the intensity of the spot is increased considerably, and a bright spot remains at that point on the screen even after the scanning spot has passed. It is possible with this type scan to produce a map of the territory surrounding the observing station on the indicator tube. This type of scan is useful when the radar set is used as an aid to navigation.

TYPE C SCAN. In the type C scan, the echo appears as a bright spot with the azimuth angle as the horizontal coordinate. This type

of scan may be used by fighters to aid in tracking an enemy plane.



Type C Scan

THE ANTENNA

An antenna is an electronic device that is used either for radiating electromagnetic energy into space or for collecting electromagnetic energy. In the radar transmitter, the magnetron generates the high frequency signal, but the antenna is needed to change this signal to electromagnetic fields which are suitable for radiating into space. The radar receiver will amplify any signal that appears at its input terminals, but an antenna is required to intercept the electromagnetic fields and to change these fields into a voltage which the receiver can interpret.

Fortunately separate antennas seldom are required for transmitting and receiving radio energy, for any antenna transfers energy from space to its input terminals with the same efficiency with which it transfers energy from the output terminals into space, assuming that the frequency is the same. This property of interchangeability of the same antenna for transmitting and receiving operations is known as antenna reciprocity. Antenna reciprocity is possible chiefly because antenna characteristics are essentially the same regardless of whether an antenna is sending or receiving electromagnetic energy. Because of antenna reciprocity, most radar sets installed in aircraft employ the same antenna both for receiving and transmitting. An automatic switch in the radio frequency line first connects the single antenna to the transmitter, then to the receiver, depending upon the sequence of operation.

Directional Properties

One of the most important characteristics of a radar antenna is its directional property or simply its directivity. Directivity means that an antenna radiates more energy in one direction than in another. For that matter, all antennas are directional; some slightly, others almost entirely. In radar operations, some antennas are required to send all energy in the same direction in order that as much as possible of the electromagnetic energy generated by the transmitter will strike an object in a given direction. In other systems, it is desirable for the energy to be radiated equally in all directions from the source.

An example of an antenna system in radar which radiates energy in a given direction is the airborne navigation and bombing set. In this set, there is only a limited amount of power available at the transmitter. In order to achieve maximum benefit from this minimum power, all of it is sent in the same direction. Since the antenna in this set is also used for reception, it also receives energy only from one direction. Because of designed features, it is possible to tell the direction of an object at which this directional type antenna is sending energy or the direction of the object from which the antenna is receiving energy. Furthermore, the physical position of the antenna is indicative of the direction of the object.

An example of a nondirectional radar antenna is the antenna installed in a radar beacon. This antenna must radiate energy equally well in all directions in order that a radar-equipped aircraft can ascertain its position regardless of its direction from the beacon antenna.

Reflectors

The two types of reflector antennas that are adaptable to fighter and fighter-interceptor aircraft are the horn and the parabolic antennas. The horn type antenna is a unidirectional antenna which may be used in day fighter aircraft. The parabolic antenna is a directional antenna which is used to sweep the area in front of the aircraft during the search phase of operation.

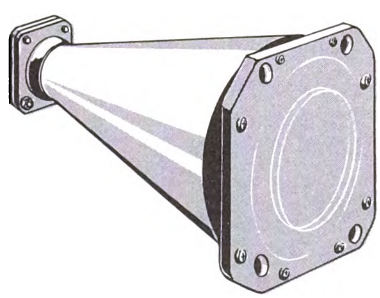
POWER SUPPLIES

All radar system vacuum tubes require voltages of various values for their filaments, screens, and plate circuits. Except for the filaments which can be heated by either DC or AC, these voltages are DC. The most convenient source of DC voltage for vacuum tubes is a rectifier. A rectifier is a device which changes AC into DC by permitting current either to flow in only one direction or to flow more readily in one direction than in the other. Devices commonly used for changing AC into DC for use in radar circuits are vacuum tube rectifiers, metallic oxide rectifiers, and crystal rectifiers.

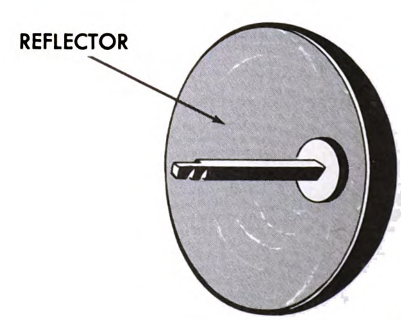
D.C. Regulated Supplies

As operation of radar circuits requires very stable voltages for optimum results, these voltages must be regulated. The circuit illustrated shows a simple regulator circuit.

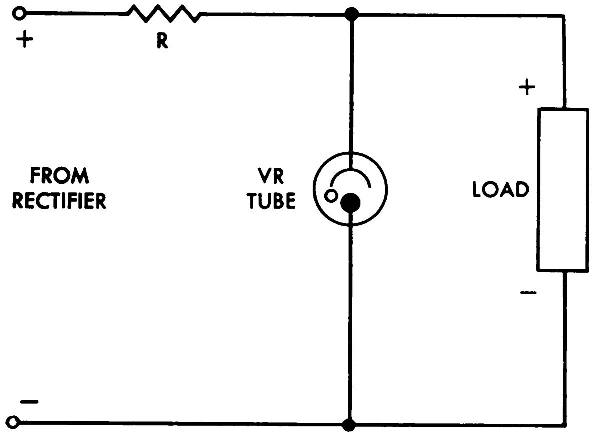
Notice that it uses a *VR* tube. Remember, too, that the *VR* tube is characterized by the fact that over a wide range of current flow there is no change in the voltage drop across it. In operation, a *VR* 150-30 tube draws current up to 30 ma and maintains the output voltage at 10 volts. If the input to the regulator were to rise, more current would flow through the *VR* tube and there would be a greater voltage drop across the resistor *R*. This voltage drop would subtract from the rising voltage and tend to maintain the voltage across the load and across the *VR* tube constant. On the other hand, if the input voltage falls, the *VR* tube will conduct



HORN TYPE ANTENNA



PARABOLIC TYPE ANTENNA



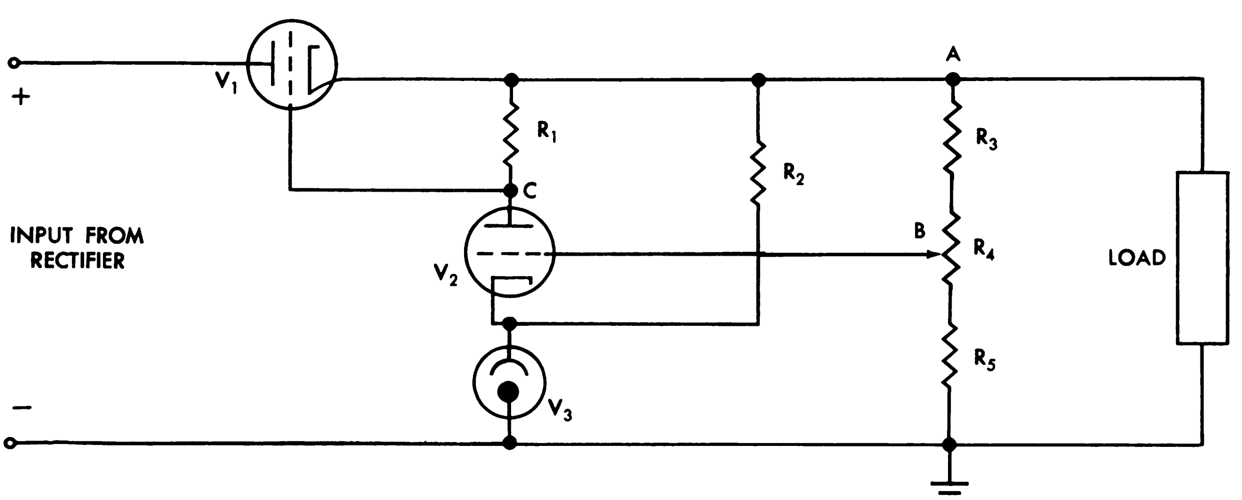
Simple Voltage Regulator

less current, the drop across R will be less, and the voltage across the load will remain at the desired level. When the load draws more current, the voltage tends to drop across R. The VR tube compensates for this by conducting less, decreasing the total current, and, as a result, maintaining the load voltage constant. Regulation with this is satisfactory as long as the current conducted remains between 5 and 30 milliamperes.

Electronic Voltage Regulator Circuit

The electronic voltage regulator shown is capable of very close regulation at a level which may be set by varying the potentiometer setting. This circuit contains a vacuum tube (V_1) in series with the load. The voltage across the load is regulated by controlling the conduction of V_1—that is, making V_1 act as a variable resistor that automatically adjusts itself to the correct value. V_3 is a VR tube, the purpose of which is to maintain the cathode of V_2 at a fixed position potential. The voltage divider system, composed of R_3, R_4 and R_5, is arranged so that the variable arm of R_4 can be adjusted to a positive voltage sufficiently low to bias V_2 for operation in the linear portion of the Eg -Ip curve. R_1 is the load resistor of V_2 and is connected in series with the VR tube, V_3. The purpose of the resistor is to absorb any changes of the voltages so that the cathode of V_2 will remain at a fixed potential.

If the output voltage (point A) tends to rise due to an increase in the input voltage or to a decrease in current drawn by the load, the voltage at point B will rise by any amount dependent upon the resistance from A to ground and from B to ground. The voltage at B is applied to the grid of V_2 and determines how much V_2 will conduct. When V_2 conducts more, there is an increase in current through the load resistor R_1 and consequently a fall in voltage at point C. This voltage is applied to the grid of V_1, and



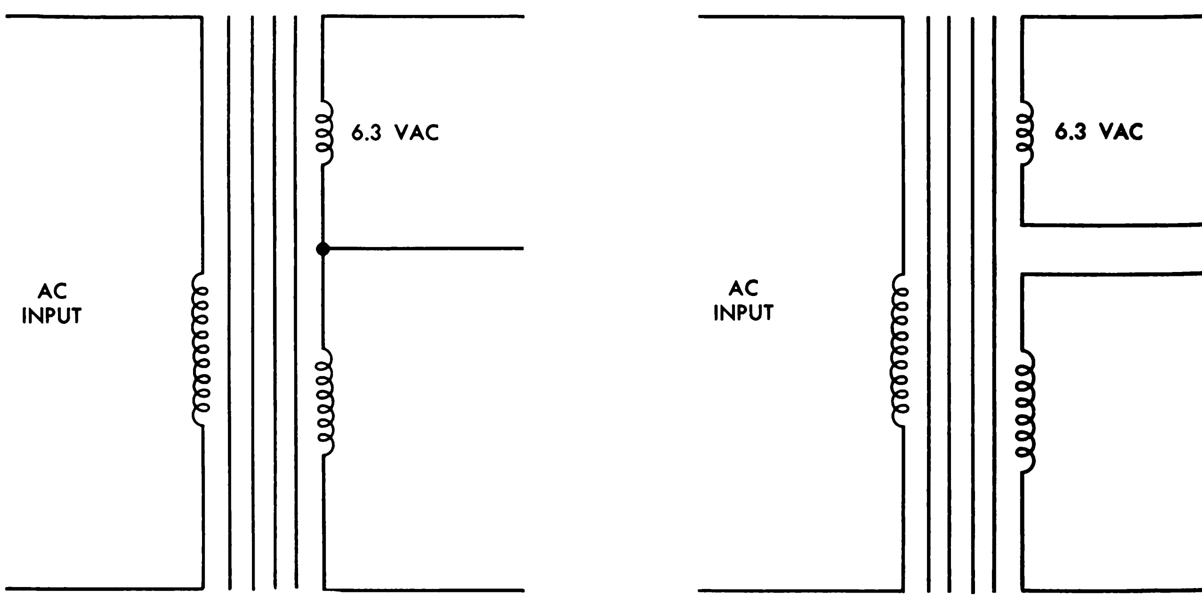
Electronic Voltage Regulator

causes it to conduct less, conteracting the tendency of the rise in output voltage which set off this action. If the voltage should tend to fall, the voltage at B drops, causing a rise in the voltage at C, which causes V_1 to conduct more, resulting in a greater output voltage. It has already been pointed out that V_2 must be biased to operate in the linear portion of the characteristic curve. This is also true of V_1, the bias of which is the IR drop across R_1.

The setting of B determines the bias of V_2 and the consequent current flow. This current flow which passes through R_1 determines the bias on V_1. The current flow through V_1 determines the load voltage. Therefore, the output voltage is adjustable, within limits, by setting the movable tap, point B, of potentiometer R_4. This circuit, with some refinements, is used in the power sources of several radar systems. Due to the great amount of current flow required in some systems, you may find two or more tubes in parallel to serve the purpose of V_1.

A.C. Power

It was previously mentioned that AC power was used to heat the vacuum tube filaments in radar system. The most common method used to obtain this voltage is from the power transformer. As an example, if 6.3 volts are needed in the circuit, it is tapped off the secondary windings of the power transformer. Another method used is a separate winding in the transformer.



Tapped Transformer—Separate Wound Transformer

CHAPTER 7

RADAR COMPONENTS OF A TYPICAL WCS

It is impossible to provide an analysis of all the possible circuit combinations to be found in the various types of radar systems. However, it is believed that the choice of circuit components presented here are sufficiently representative of all Weapons Control Systems to contribute materially toward the development of a more intelligent servicing technique.

The five basic functions of a Weapons Control System are introduced in this chapter. Each function is then broken into the various components comprising it.

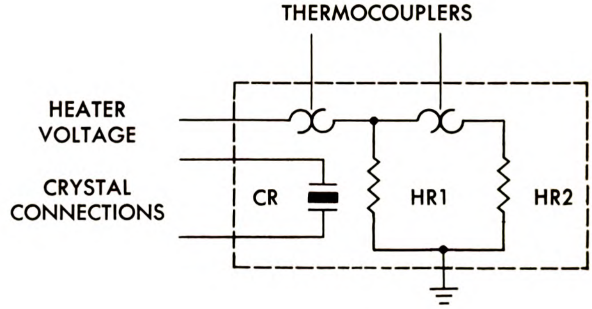
TRANSMITTER FUNCTION

The transmitter function produces high-powered X-band pulses of RF energy. These pulses are radiated into space for the illumination of airborne targets.

Crystal Oven

When an interceptor takes off at sea-level and climbs to an altitude of 40,000 feet, a vast change in temperature is encountered. This vast change of temperature has such an effect on a crystal that its resonant frequency is changed. To prevent this change in frequency the crystal is installed in an oven where the temperature is held nearly constant by heater resistors under the control of a coarse and a fine thermocoupler.

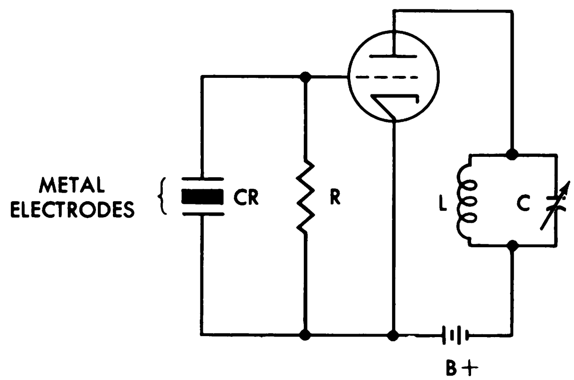
Referring to the illustration of a *Crystal Oven,* it can be seen that everything is sealed in an airtight container. Assuming that we want a constant temperature of 78°F., heater voltage would be felt across the thermocouplers and the heater resistors. The current through the heaters will cause the temperature to rise. When the temperature is within approximately 15° of that desired, the coarse thermocoupler contacts open the circuit to *HR2*. *HR1* will continue to have current flowing until the temperature reaches 78°F., then the fine thermocoupler opens and remains open until the temperature has dropped approximately 2° to 5° below the desired temperature. At this time, the fine thermocoupler will again close, allowing current to flow through *HR1* until the desired temperature is again obtained.



Crystal Oven

Crystal Oscillator

A quartz crystal can be used to control the frequency of an oscillator if it is so placed in the circuit that it takes the place of the normal frequency controlling circuit. This is shown in the illustration of the *Crystal Oscillator.*

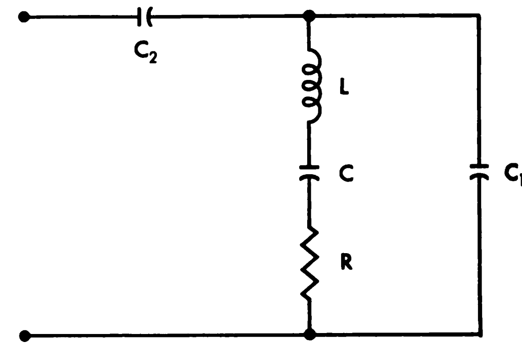


Crystal Oscillator

The current which a crystal develops at its natural or resonant frequency is the same as that produced by a series resonant circuit consisting of resistance, capacitance, and inductance.

When an alternating voltage is applied across a crystal in a crystal oscillator circuit, the crystal vibrates. If the frequency of the voltage is approximately equal to the natural vibrating frequency of the crystal, the crystal produces a high amplitude voltage at this frequency.

Because of the property of displaying series resonance when connected in a circuit, the crystal can be represented by the *Equivalent Circuit of a Crystal.* In the illustration C_1 represents the capacity between the two electrodes. The crystal acts as the dielectric between the two electrodes. When the crystal is placed in a vacuum tube circuit, the capacity changes to a value depending upon the crystal holder, the input capacity of the oscillator tube, and the capacity of the connecting wires between the crystal and the tube. The inductance L represents the crystal mass, C the resilience (the ability of the crystal to spring back to its original shape after being distorted by an applied voltage), and R frictional losses. The component C_2 represents the series capacity between the crystal and its electrodes.



Equivalent Circuit of Crystal

The frequency at which the reactances of L and C produce series resonance is the natural vibrating frequency of the crystal. When a voltage is applied across the crystal at a slightly higher frequency than its natural frequency, the effective reactance of L and C combined becomes inductive and numerically equal to the reactance of C. At this frequency, antiresonance occurs, and the crystal becomes the equivalent of an antiresonant or a parallel resonant circuit.

C_2 affects the circuit only when the crystal electrodes do not make close contact with the faces of the crystal. A decrease in its value produces an increase in the resonant frequency.

Binary Counter

A binary counter is basically an Eccles-Jordan Multivibrator (refer to chapter 4, Section F) which takes a basic frequency and divides it by two. Therefore, the output frequency of a binary counter would be half the input frequency. This is a very desirable feature due to the fact a crystal is commercially cut to resonate at the frequency of 356 kc. To have a crystal cut to a speci-

fied frequency would result in greater cost than the use of several binary counters in series.

Blocking Oscillator

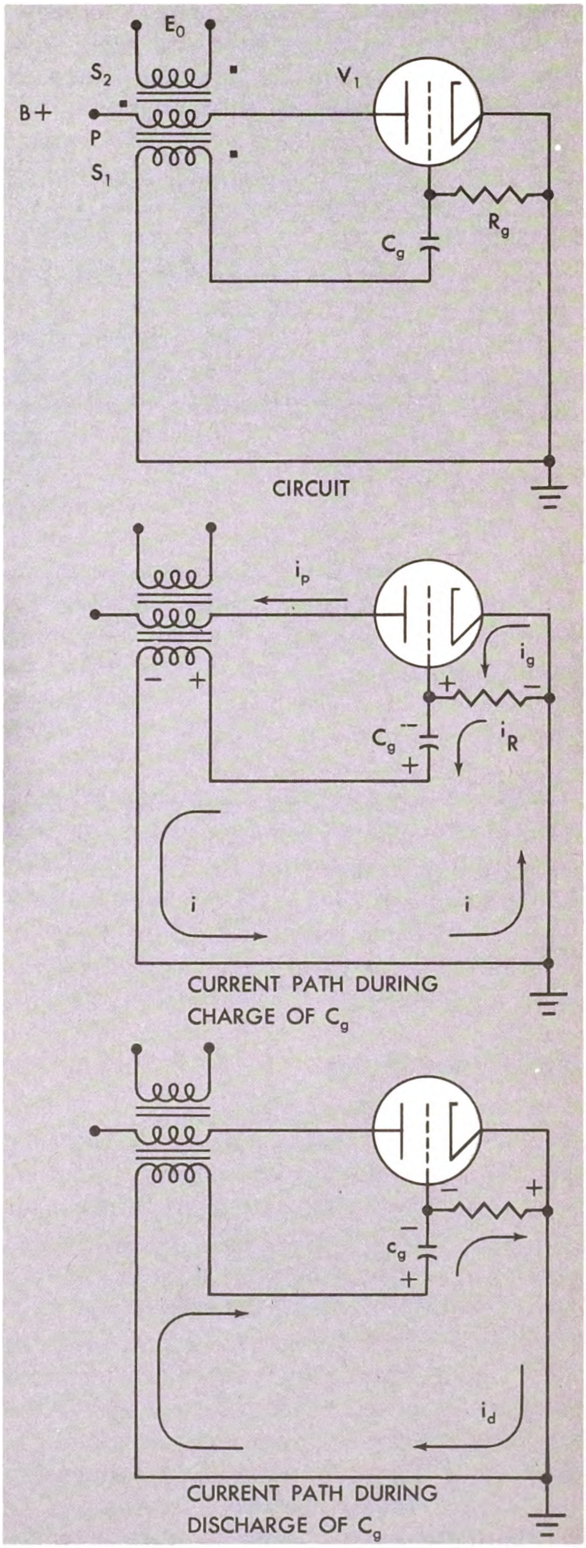
The blocking oscillator is a type of relaxation oscillator. It can do many of the things that a multivibrator does. Thus, it can be free-running, synchronized at each cycle or at a submultiple frequency, or it can be driven in the manner of a one-shot multivibrator.

A blocking oscillator is any oscillator which cuts itself off after one or more cycles because of the accumulation of a negative charge on the grid capacitor. The pulse duration is determined by the size of the grid capacitor which charges through the grid-cathode resistor of the tube during the positive swings of the oscillations. After the capacitor has become charged to the voltage at which oscillations cease, grid current ceases to flow and the capacitor discharges through the grid-leak resistor. The illustration of the *Single-Swing Blocking Oscillator* shows the charge and discharge paths of the capacitor. The value of the grid-leak resistor establishes the pulse-repetition frequency.

Any small changes in the circuit constants or operating conditions will affect the stability of the pulse rate. Where it becomes desirable to maintain a pulse-repetition frequency at a value more nearly constant than that which is possible by grid-blocking action above, a synchronizing signal is applied to the grid or plate of the oscillator. The grid constants are adjusted to produce the approximate repetition rate, and the exact operating frequency is established by the synchronizing voltage.

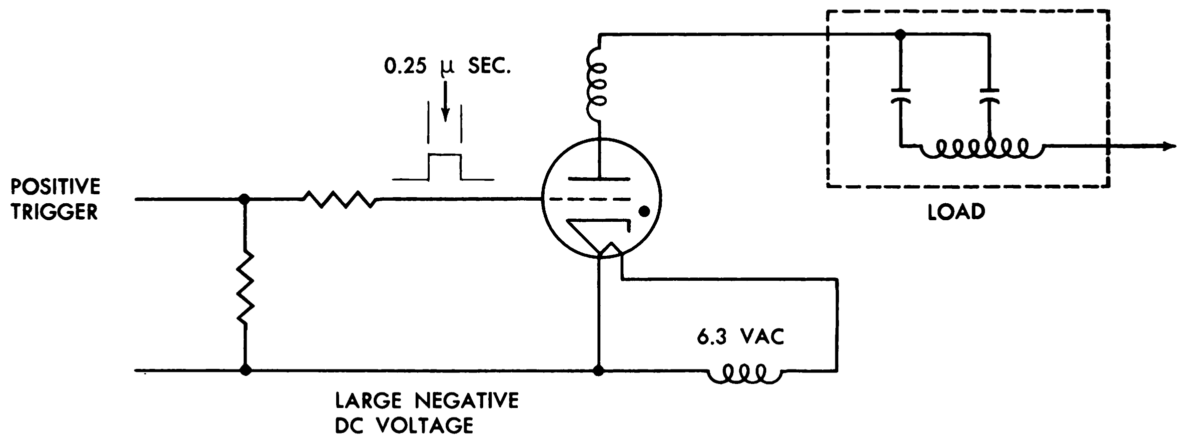
Thyratron Modulator

The thyratron is a gas-filled triode, as shown in the *Thyratron Modulator* illustration. When the thyratron is used as a modulator, a large negative voltage is generally applied to its cathode which biases the tube



Single-Swing Blocking Oscillator

below cutoff until a large enough positive voltage is felt on the control grid. This positive voltage is generally a square wave or pulse of an extremely short duration.



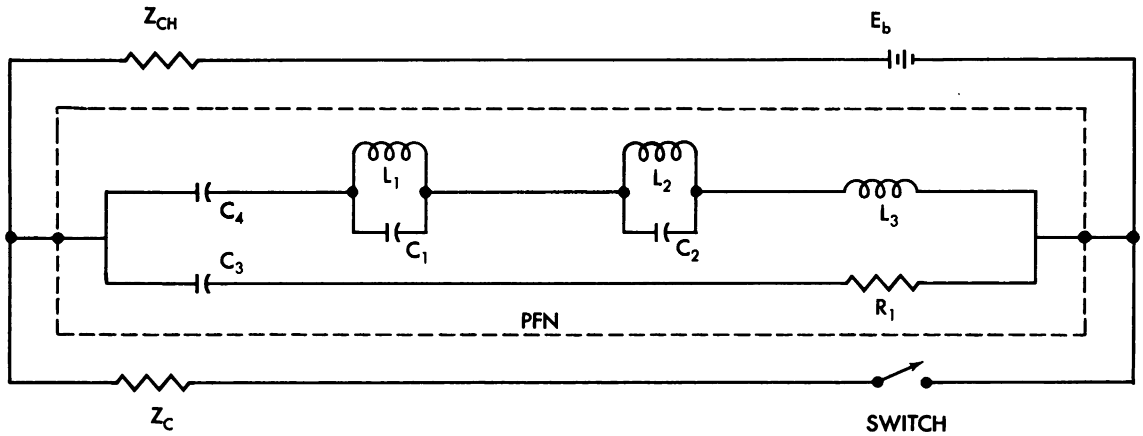
Thyratron Modulator

After the pulse is felt on the control grid causing the tube to fire and ionize the gas, the thyratron acts as a very low resistance conductor, for the plate voltage is now at the high negative voltage that was applied to the cathode. So actually the thyratron modulator is a highly effective electronic switch.

Pulse Forming Network

The pulse forming network serves the double purpose of storing energy and of shaping the output waveform. In operation it is quite similar to the artificial transmission line; that is, it furnishes a steady output voltage for a certain length of time, at the end of which time its voltage drops to zero. It is different, however, in that instead of being made up of a number of identical sections, each section is different. It has the advantage of forming more nearly rectangular pulses than an artificial transmission line with a similar number of sections.

Although the theory of operation of the *Pulse Forming Network* shown is somewhat complex during discharge, it may be summed up as follows: C_4 stores most of the energy for the pulse while the components C_1, C_2, L_1, L_2 and L_3, which are in series with it, mainly shape the pulse. During the first part of the pulse (when the charge on C_4 is nearly equal to the supply voltage E_b) the series components offer a high impedance and cause the voltage across Z_c (the load) to jump to $E_b/2$. As the pulse continues, C_4 discharges, but the impedance of the series components drops also. This causes the voltage across Z_c to remain at $E_b/2$. Later, when the charge on C_4 is less than $E_b/2$, the inductances maintain the level of current flow. This continues until the condenser is discharged and the magnetic fields are collapsed. At this time the voltage drops rather rapidly to zero. The components C_3 and R_1 make the leading edge of the pulse steeper. There being no inductance in this branch of the network, the current can rise immediately to its full value. The pulse which appears at the load is essentially a rectangular pulse of the desired duration and equal in amplitude to $E_b/2$. The values of the components determine the length of the pulse and the characteristic impedance of the line. The characteristic impedance of the line must be matched to



Pulse Forming Network

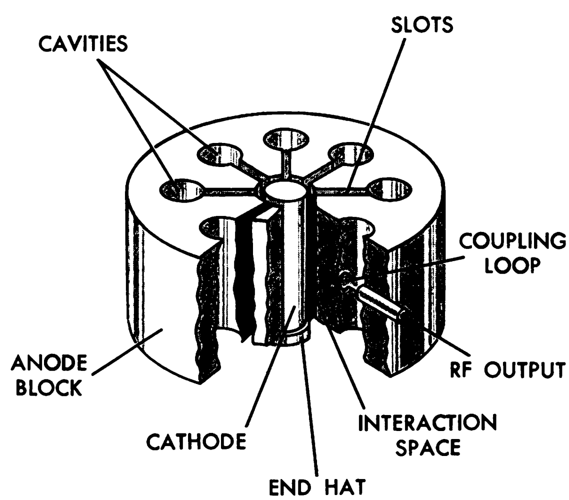
that of the load. This accounts for the fact that even though the line is charged to E_b, the voltage across Z_c is only $E_b/2$.

Tunable Magnetron

The magnetron oscillator is the only available source of large amounts of RF power, either CW or pulsed, at frequencies exceeding 1,000 megacycles. Microwaves of the order of 1 cm and peak power outputs as high as 5,000 kilowatts have been generated with magnetrons. As much as 1 kilowatt of CW power has been obtained from magnetrons operating with wavelengths of 10 cm.

The multicavity or cavity magnetron is the only type in use today. The internal structure is shown in the illustration of a *Multicavity Magnetron* structure. The anode is a block of copper into which an even number of cavities have been cut. The number of cavities may vary from 6 to 18; each cavity is connected to the central chamber by a slot or aperture. The dimensions of each cavity and slot are so chosen that their resonant frequency in the lowest mode determines the output frequency. Frequency is also affected to a lesser degree by the end space and axial length of the anode block.

Since the various cavities are coupled together, a coupling loop placed within one effectively abstracts energy from all cavities. The whole structure is enclosed in an evacuated metal envelope. The magnetron requires an external magnetic field so oriented that the flux lines are parallel to the axis of the cathode. This field is usually provided by a permanent magnet which can be separated from the magnetron. If the magnetic pole faces are built into the magnetron structure itself and the magnet attached to it, the magnetic field gap and total magnet weight are reduced. This design has resulted in the



Multicavity Magnetron Structure

so-called "packaged magnetrons" used at shorter wavelengths, where the magnetic fields are high but need not be spread over a large area.

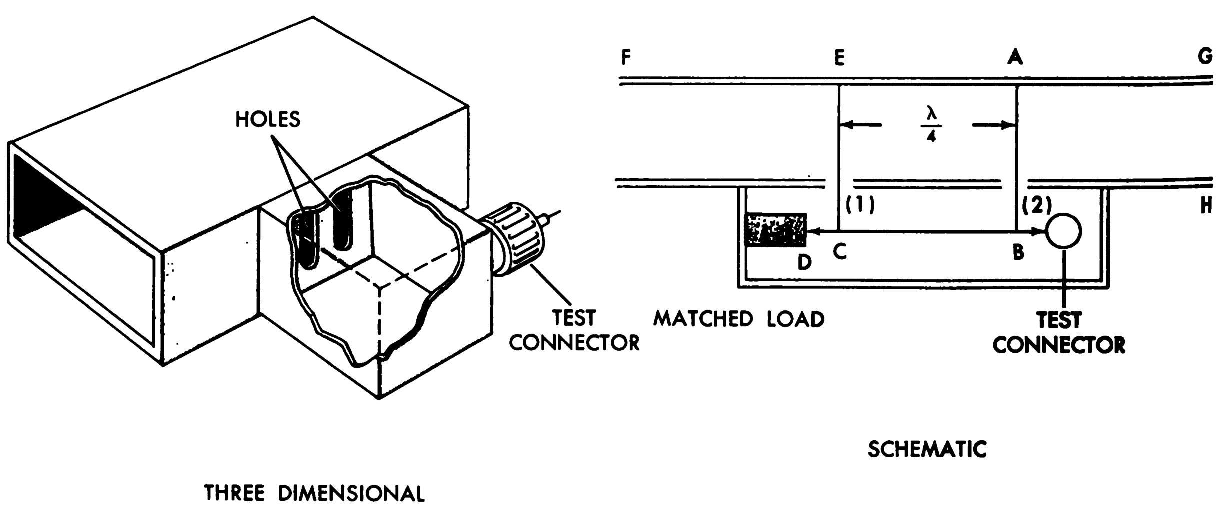
Magnetrons that are tunable have also been manufactured. Four principal methods are used for tuning magnetrons. One method consists of employing a ring between the straps which adds capacitance and lowers the resonant frequency by an amount which depends upon the position of the ring. A second method uses a ring placed over the ends of the cavities and raises the resonant frequency by lowering the inductance as the ring is brought closer to the anode block. A third method uses cylindrical conducting rods inserted in the cavities parallel to the cathode. These plugs result in decreased inductance, thus increasing the resonant frequency. A fourth method employs the coupling of a separate tunable cavity to one of the anode cavities. Because this cavity is closely coupled to one resonant cavity and all the resonant cavities are closely coupled to each other, it is possible to control the resonant frequency of the magnetron as much as six percent.

Directional Coupler

The directional coupler provides a convenient means of measuring transmitted power and receiver sensitivity. When the coupler is in use, it is located in series with the waveguide between the transmitter and the antenna.

The illustration of the *Directional Coupler* shows that the unit consists of two parallel sections of waveguide welded along the narrow edge with two circular holes in the common wall. These holes are located one quarter wavelength apart and couple a small amount of energy from the main waveguide into the parallel guide in which a resistance strip is located. This strip is a waveguide characteristic impedance matching device and causes the guide to appear electrically as an infinitely long waveguide. At the other end of the attached section is a test connector to which the test equipment may be attached.

The directional coupling properties of the coupler are obtained as follows: A part of the energy coming from the transmitter unit and flowing toward the antenna will be coupled through holes (*1*) and (*2*). Equal



Directional Coupler

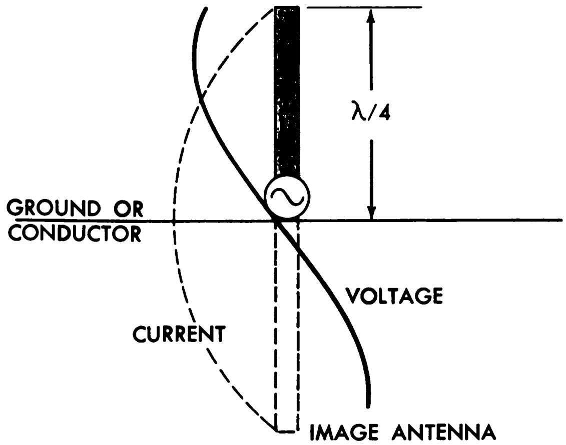
amounts of energy will follow the path *FEABH* and the path *FECBH*. Since the energy traveling along one of these paths travels exactly the same distance as the energy flowing along the other path, the RF energy flowing along each path will be in-phase and additive at the test probe. However, energy following the path *FEABCD* will have traveled one half wavelength further than that following the path *FECD* and will energize the test probe with little net energy. Thus, a very small amount of the energy traveling toward the antenna in the main waveguide is coupled into the matching load, and all the coupled energy in the attached waveguide section is applied to the test connection. Similarly, energy flowing along paths, *GABH*, *GAECBH*, and *GAECD* is either dissipated at the matched load or out of phase and subtractive at the test probe. Therefore, the output at the test probe will be a measure of the transmitted power and the accuracy of this measurement will depend upon the selectivity of the coupler. The directional coupler has a nominal "directivity" of 20DB, and each unit is calibrated and marked.

Antenna Devices

The three most commonly known types of antenna are the Hertz, the Marconi, and the folded Dipole.

The Hertz antenna, a half wavelength conductor, is the simplest of the radiating elements. Considerable radiation occurs in this element because of its resonant characteristics and its ability to store large amounts of energy in induction fields. Resonance causes high voltages and high circulating currents, and they, in turn, produce strong fields around the antenna. An antenna of the correct length acts like a resonant circuit and presents pure resistance to the excitation circuit. An antenna having other than the correct length displays both resistance and reactance to the excitation circuit. An antenna slightly longer than a half wave, for example, acts like an inductive circuit.

The Marconi antenna is a quarter wavelength long with one end grounded. The grounded end then acts as a reflection of a quarter wavelength so effectively we have a half wavelength antenna as shown in the illustration of the *Marconi Antenna.* The grounded end or conducting plane is usually the skin of the aircraft where airborne equipment is concerned.



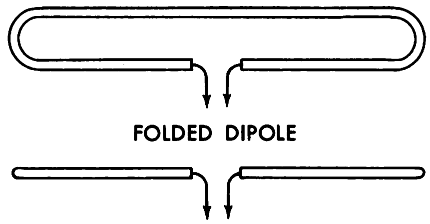
Marconi Antenna

The folded dipole antenna is the one most often employed in radar equipment. It is a full wavelength conductor which is folded to form a halfwave element as shown in the illustration of a *Folded Dipole.* A better description is that it consists of a pair of half-wave elements connected together at the ends. In it the voltage at the ends of each element must be the same. With the same amount of current induced in each element, the field strength in space is doubled. This causes the power density per square meter to increase four times.

TARGET DETECTION AND DISPLAY FUNCTION

This function is performed by sections which provide detection and amplification of weak echo signals that are returned to the radar receiver when RF energy radiated by the antenna strikes the target. The sections

convert the echoes into useful information for presentation on the indicator in the form of a visible display.



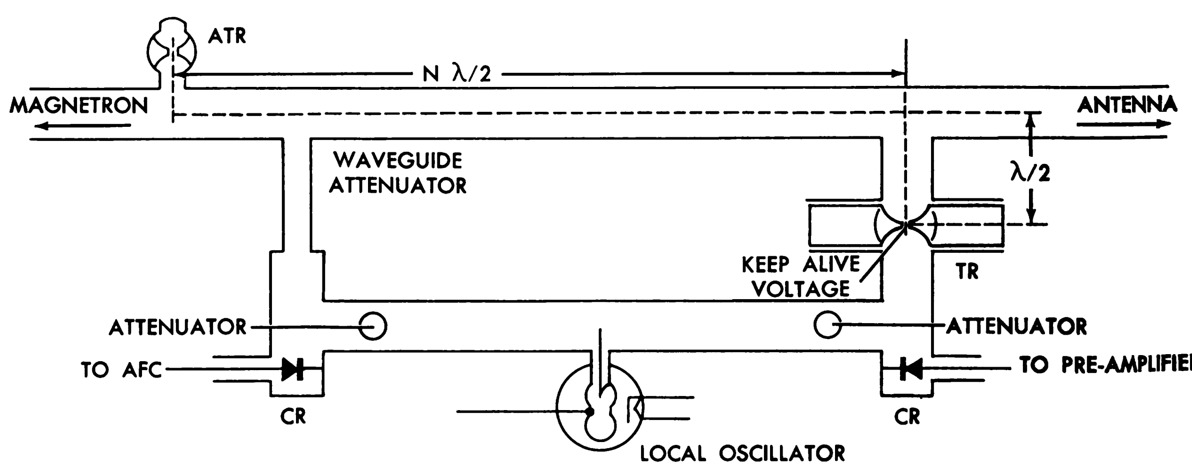
Folded Dipole

Duplexer

The pulse-modulated RF output signal from the magnetron oscillator and the received echo pulse pass through the antenna and duplexer. The duplexer together with the transmit-receive (TR) tube and the anti-transmit-receive (ATR) tube is essentially a two-position waveguide switch which prevents the transmitter high-power pulse from entering the receiver crystal channel and thereby damaging the receiver crystal. It also keeps a portion of the echo signal from entering the magnetron and being dissipated there.

The TR tube is mounted on the waveguide assembly so that it is in series with a branch from the main waveguide to return incoming signals (echoes) to the receiver. Reference should be made to the illustration of the *Duplexer*. The tube consists of a resonant cavity with a spark gap at its center filled with a low pressure gas. The received signals, being at a low-power level, will not cause an arc discharge across the TR electrodes and thus will pass through the TR with only slight attenuation. On transmission, however, the gap breaks down and short-circuits the cavity, thus preventing the high power delivered by the magnetron from damaging the highly sensitive crystal. Inside one of the gap electrodes is an auxiliary "keep-alive" electrode, which has approximately a —1,000 volt DC applied to it to maintain initial ionization of the gas in the vicinity of the spark gap. This enables rapid ionization of the gas.

The ATR tube is mounted between the TR tube and the magnetron. The tube contains a resonant cavity and is so designed as to present a high impedance in series with the surface of the waveguide when de-ionized. This high impedance plane is an integral number of electrical half wavelengths from the TR section. As a result, received signals coming back through the waveguide will



Duplexer

find a high impedance toward the magnetron and, therefore, will be diverted through the TR tube into the receiver mixer. During transmission, the magnetron RF pulse causes the ATR tube to break down so that it has a low impedance and allows the magnetron power to travel to the antenna. The ATR tube consists of a piece of waveguide filled with low-pressure gas, closed by a glass window at one end.

Transitron Oscillator

The transitron is a negative resistance oscillator. A resonant circuit, we have found, has a definite value of parallel impedance at resonance. If we connect a negative impedance into the circuit equal to the parallel impedance, the two will cancel out giving us, in effect, a circuit containing no impedance.

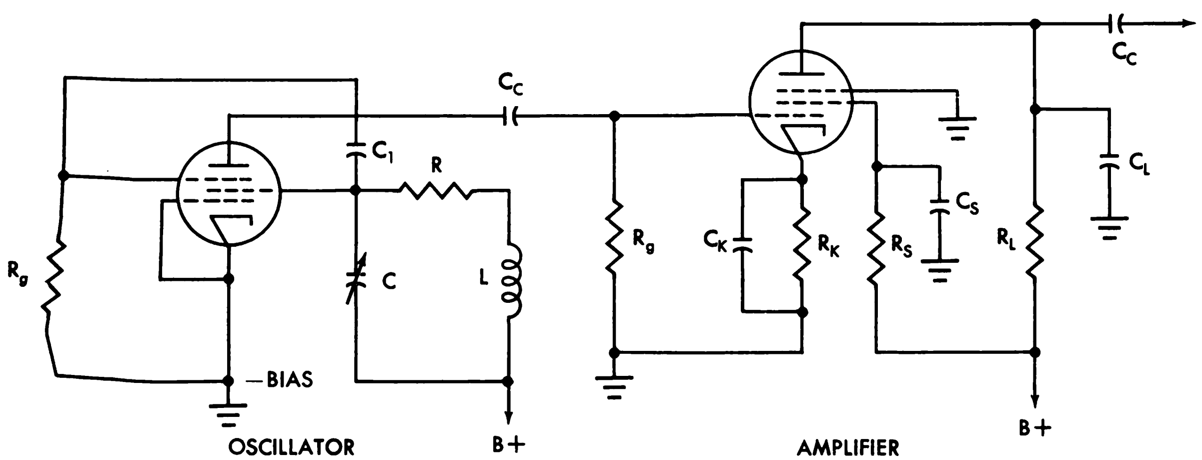
To produce this negative resistance we place a negative bias on the suppressor grid of a pentode as shown in the *Transitron Oscillator* illustration. This negative bias then turns back, to the screen grid, electrons that normally would reach the plate. Thus, we have increased the screen current and reversed the normal tube action. In doing this we have not consumed any power, but have generated it. The signal is then taken and passed through a coupling capacitor to the amplifier. There the signal is amplified and sent to the next stage of the radar circuit.

Discriminator

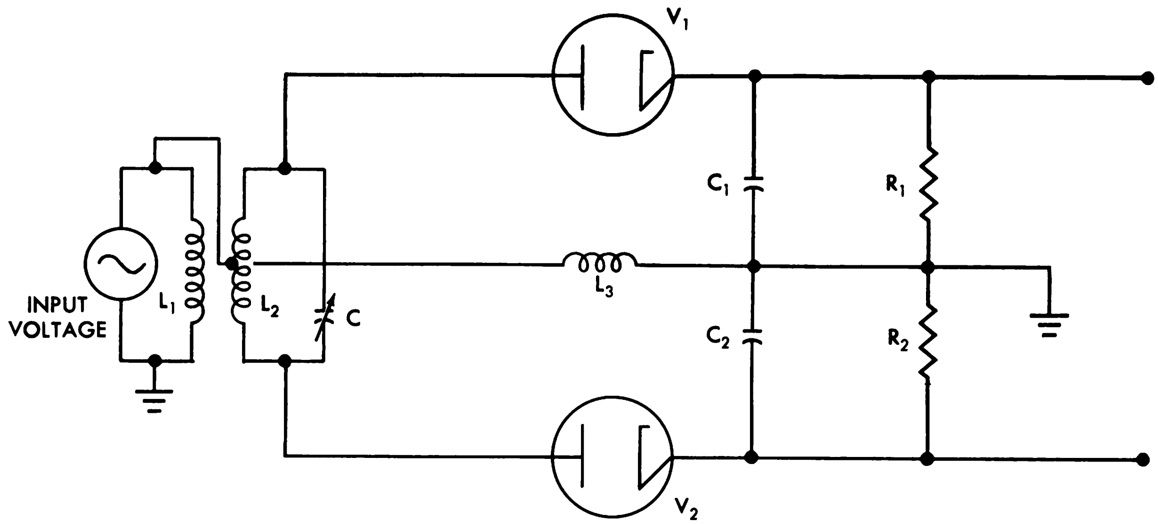
The frequencies of the transmitter occasionally change, causing considerable trouble at the receiver. The receiver is no longer tuned exactly to the RF signal frequency, and signals which supply weak returns may no longer be visible on the indicator screen. Therefore, most radar receivers employ an automatic frequency control circuit in which changes of frequency are detected by a discriminator circuit.

The detection of a change of frequency by the discriminator is accomplished by a transformer, which, due to special connections, distorts the frequency change into a voltage of varying amplitude change. After distortion, the voltage of varying amplitude is detected in the manner employed by any amplitude-modulation detector. The illustration shows the circuit of a typical *Discriminator Circuit.*

The two coils L_1 and L_2 are the primary and secondary, respectively, of the IF transformer. The secondary coil is tuned to the correct transmitter frequency by condenser



Transitron Oscillator



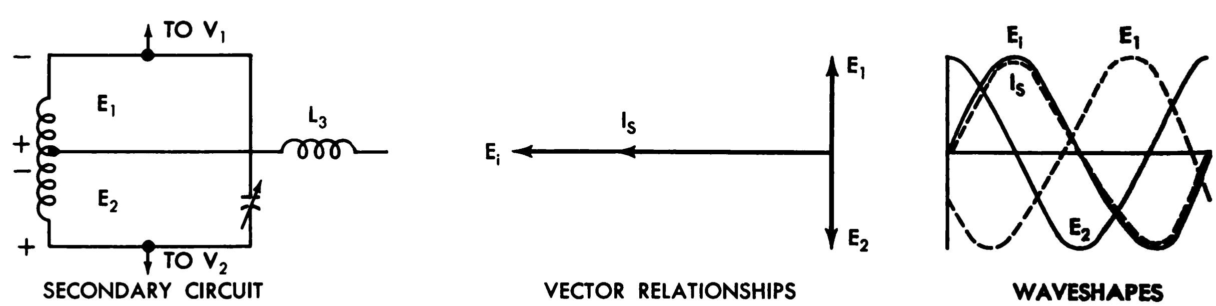
Discriminator Circuit

C. Tubes V_1 and V_2 are diode detector tubes. R_1 C_1 and R_2 C_2 are filters for the removal of the RF component from the circuit.

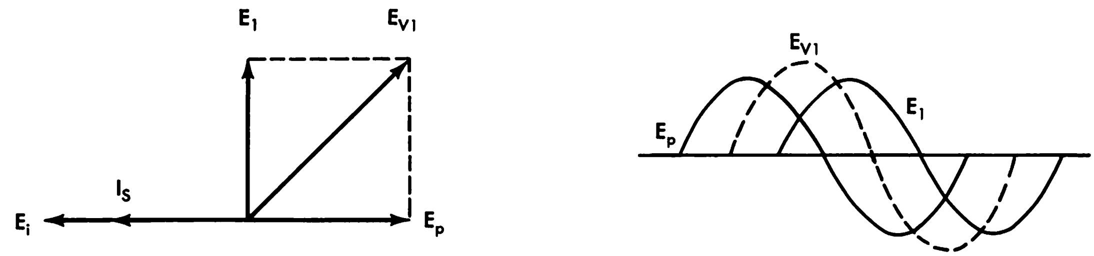
L_3 is an RF choke which has a high reactance to the RF energy. Since it is connected across the primary, the primary voltage appears across it at all times. The connection to each diode from this choke causes the primary voltage to appear at the plates of the diodes with the same phase in each case. This voltage causes currents to flow in opposite directions in the resistors R_1 and R_2, resulting in a zero output. Therefore, the discriminator is not affected when the amplitude of the applied voltage changes.

Since the circuit is at resonance, the secondary current (I_S) is in phase with the induced voltage (E_i). Each half of the secondary has considerable inductive reactance at the RF frequency; therefore, there is a voltage drop across this reactance due to the secondary current. The voltages in each half of the secondary winding are 180° out of phase with each other due to the manner in which the secondary winding is connected to the tubes. Current flows in the same direction through both coils, but the vector voltages ($E_1 + E_2$) are measured with respect to the center. This causes opposite polarities to exist. The illustration *Secondary Voltages at Resonance* shows this.

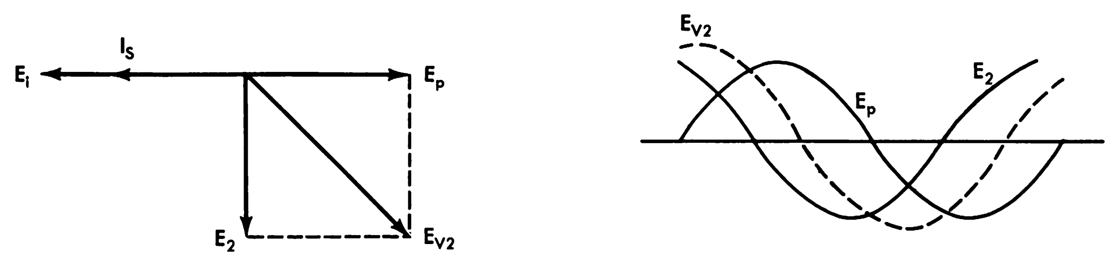
The plate voltage on the plate of V_1 is the sum of the primary voltage (E_p) and the voltage across the upper half of the sec-



Secondary Voltages at Resonance



TUBE V_1



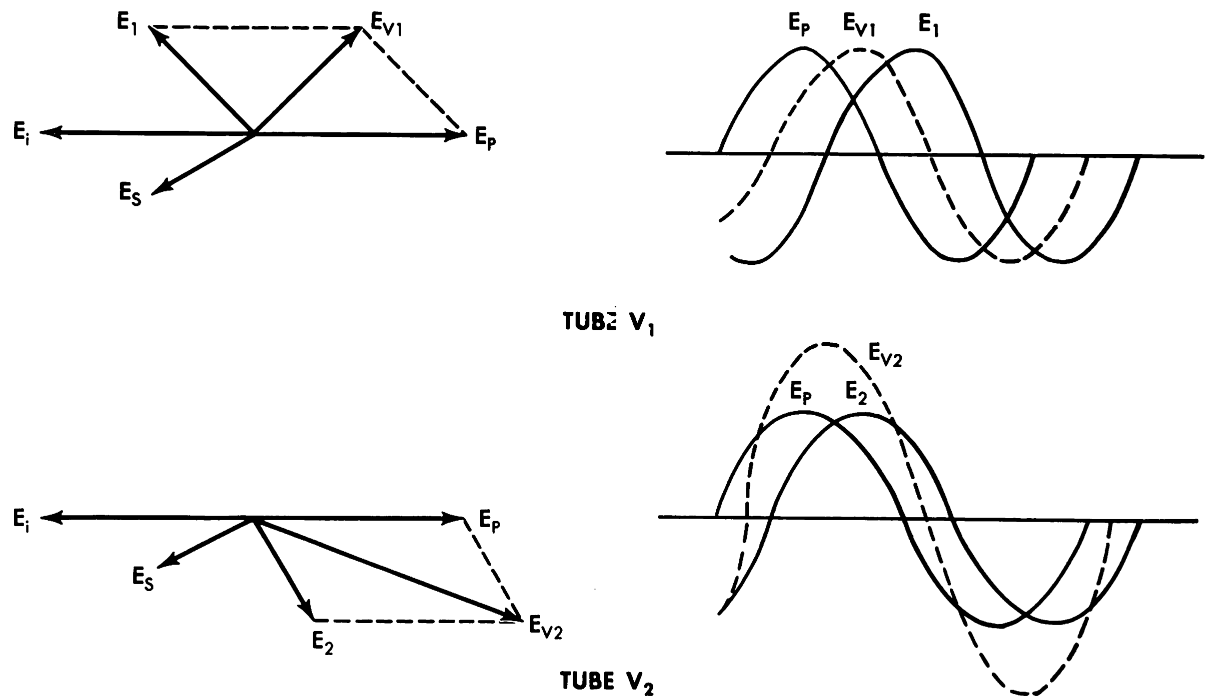
TUBE V_2

Resultant Plate Voltages at Resonance

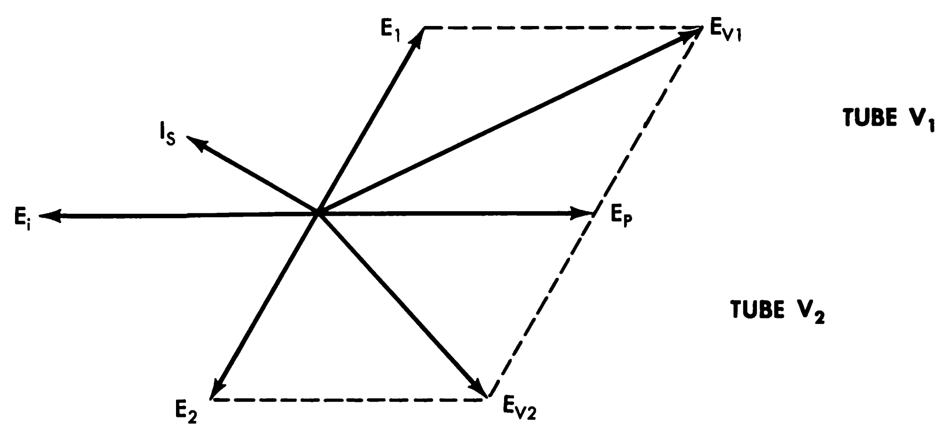
ondary (E_1). The plate voltage of V_2 is E_p and E_2. Because these voltages are out of phase with each other, they must be added vectorially to obtain the actual plate voltages, E_{V1} and E_{V2}, as shown in the illustration of *Resultant Plate Voltages at Resonance.* When the plate current flows, C_1 and C_2 charge to near the peak voltage of E_{V1} and E_{V2}, producing DC voltages across R_1 and R_2. Note that these are DC voltages and that phase relationships do not exist at this point. Since E_p is common to both tubes and E_1 equals E_2, the DC voltages are alike but of opposite polarity; therefore, their sum and the output will be zero.

When the frequency is lower than the resonant frequency, the output will no longer be zero. Since inductive reactance decreases as frequency decreases, while capacitive reactance increases, the secondary circuit of the transformer will be capacitive. This means the secondary current I_s will now lead the induced voltage (E_i); but the voltage drops E_1 and E_2 will still be 90° out of phase with the current, since this is not a function of frequency. These conditions are shown in the illustration of *Voltages at Frequencies Below Resonance.* Under these conditions, E_1 is now more than 90° away from E_p when added vectorially; this makes their sum less at resonance even though the magnitudes of E_p and E_1 did not change. At the same time E_2 is less than 90° away from E_p so the magnitude of the resultant E_{V2} is greater than at resonance. Again, since the resistor voltages depend upon the length of the resultant vectors, there will be a high voltage across R_2 and a lower voltage across R_1. Therefore, the output voltage will be equal to the R_2 voltage minus the R_1 voltage, and the output will be negative with respect to ground.

When the input frequency is greater than the resonant frequency, the secondary circuit is inductive, and the current I_s lags the induced voltage. The illustration of *Voltage at Frequencies Above Resonance* show that E_{V1} has a greater magnitude than E_{V2}. Therefore, the DC voltage across R_1 is



Voltages at Frequencies Below Resonance



Voltage at Frequencies Above Resonance

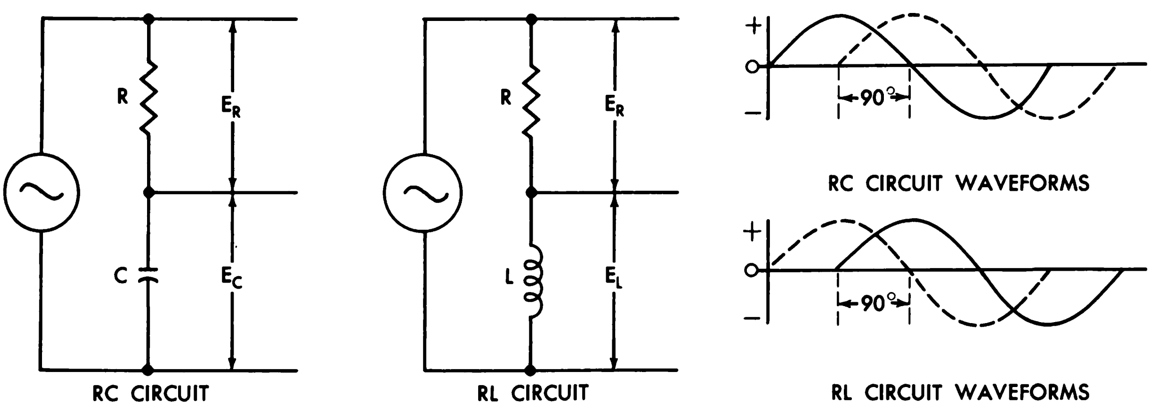
greater than the voltage across R_2, making the output positive with respect to ground.

Phase Splitter

A phase splitter is a circuit which produces, from the same input, two output waveforms which differ in phase from one another. Two types of circuits used for producing two out-of-phase output waveforms from the same input are the RC phase-splitting circuit and the LC phase-splitting circuit.

Referring to the illustration of an *RC and RL Phase Splitter* circuit, notice that *R* and

C which are connected in series represent the load impedance. The same current flows through both R and C. The voltage across R is always in phase with the current in it, since, in a resistive circuit, current and voltage are in phase. In a capacitive circuit, current and voltage are 90° out of phase, with current leading. Therefore, the voltage across C is 90° out of phase with the current. Thus, the two voltages—that across R and across C—are 90° out of phase with each other. An identical phase shift occurs in the RL circuit. The only difference is that the current lags the voltage in L by 90°. The voltages, however, like those across R and C in the RC circuit, are 90° apart.



RC and RL Phase Splitter

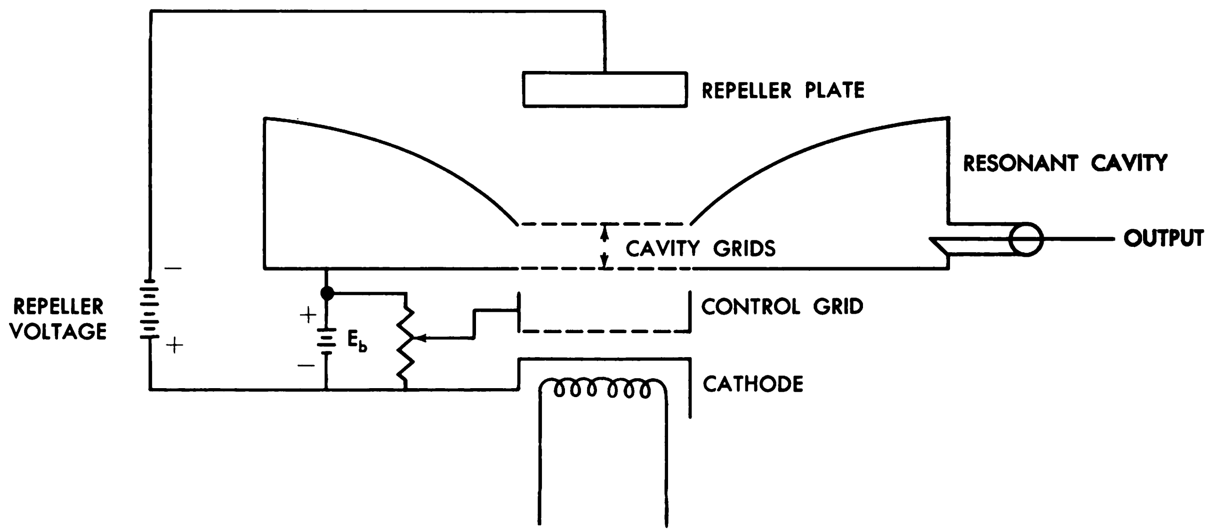
Multivibrator

Multivibrators produce a specified waveshape over and over again as long as power is supplied, or produce one wave when pulsed, remain quiescent until another pulse is applied to the circuit, and then produce another wave. Regardless of the waveshape or amplitude of the trigger pulse, the output waveshape is the same. Those which run continuously with only power applied are called free-running multivibrators. Those which must be started for each cycle are called one-shot multivibrators.

These circuits can produce sawtooth, triangular, or square waveforms, the duration and amplitude of which can be readily varied. Refer to Chapter 4 for the principles of the multivibrator.

Reflex Klystron

The reflex klystron is normally used as a local oscillator in the radar set. It is a tube of the velocity-modulated type which has only one cavity resonator and acquires its name from the action of the repeller plate which turns the electrons back to the cavity grids. The repeller plate does this because the voltage applied to it is negative with respect to the cathode. The illustration is a *Schematic Representation of a Reflex Klystron* showing the voltages applied to the electrodes and cavity grids. The tube consists of an indirectly heated cathode, a control grid which acts as a focusing electrode, a re-entrant cavity resonator built around two cavity grids, and a repeller plate. Because the potential in the cavity grids is 200-300 volts positive with respect to the cathode, the electrons emitted are attracted towards them. The control grid, which is slightly positive with respect to the cathode, focuses the electrons into a beam. After passing through the cavity, the beam of electrons enters the field between the cavity grids and the repeller plate. The repeller plate is maintained at approximately 100

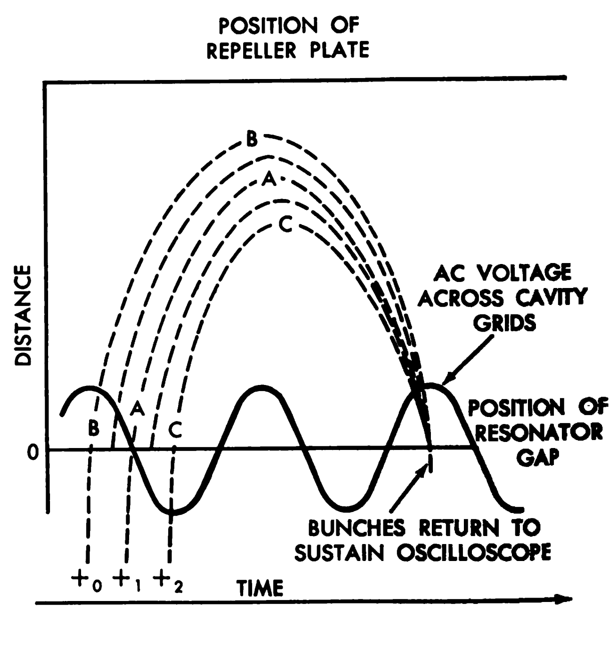


Schematic Representation of a Reflex Klystron

volts negative with respect to the cathode. The beam is therefore decelerated by a field of approximately 300-400 volts, brought to a stop, and repelled toward the cavity.

Since this tube is of the velocity-modulated type, the electrons must be bunched and then brought back to the cavity in proper phase to deliver energy to it. The illustration *Bunching Action in a Reflex Klystron* shows how this takes place. Electrons emitted by the cathode are attracted toward the cavity because of its high positive DC potential. Electrons arrive at the resonator gap when the AC voltage across it is either positive, zero, or negative. An electron (*A*) arriving at the resonator gap at time t_1, when the AC voltage is zero, moves forward without a change in velocity; hence, it may be considered to be a reference electron. This electron travels into the drift space between the repeller plate and the resonator gap until it is decelerated sufficiently to be turned back toward the cavity. Another electron (*B*), arriving at the resonator gap at time t_0, when the AC voltage is positive, is accelerated and moves forward at a higher velocity than the reference electron; hence, it penetrates farther into the drift space, before being decelerated sufficiently to be turned back, than electron *A* did.

A third electron (*C*), arriving at the resonator gap at time t_2, when the AC voltage is negative, is decelerated and moves forward at a lower velocity than the reference electron; hence, it does not penetrate as far into the drift space as the reference electron did before being decelerated sufficiently to be turned back.



Bunching Action in a Reflex Klystron

Although the three electrons travel different distances into the drift space and the velocities of the three are different, they are bunched during their travel so that they arrive back at the resonator gap together. Note that electron *B*, which starts before the reference electron and travels at a greater velocity, also has greater transient time. Similarly, electron *C*, which passes through the gap later and travels with less velocity than the reference electron, has a shorter transient time.

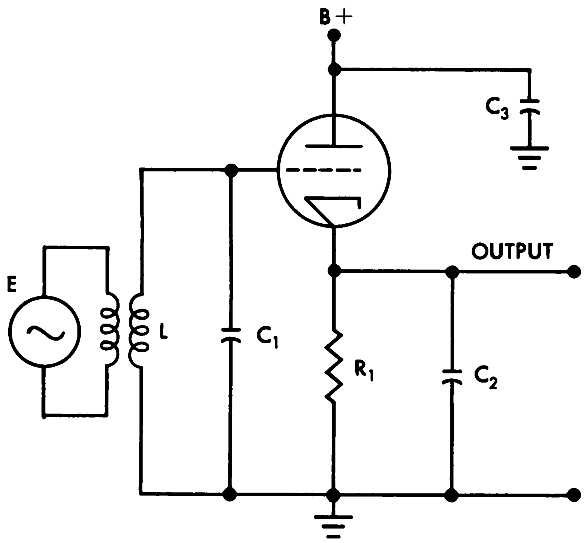
We have seen that the reflex klystron velocity-modulates the electrons so that they arrive back at the resonator gap in bunches. In order for these bunches of electrons to deliver energy to the resonator and thus sustain oscillation, they must be decelerated as they approach the grids. When the AC voltage across the cavity grids is negative, the bunches of electrons are accelerated. It is therefore evident that, for the transfer of energy from the bunched electrons to the resonator to be a maximum, the bunches should arrive at the resonator gap at a time when the AC voltage across the grids is at a positive maximum. Since a positive maximum occurs once each cycle, oscillations are sustained if the time of arrival of the bunches at the grids is made to coincide with any one of these positive maximums.

Detector

It is the function of the detector to reproduce the signal frequencies from the variations of the carrier frequency and its sidebands. There are different types of detectors, but here we shall only discuss the infinite-impedance detector.

The infinite-impedance detector has the ability to handle large input amplitudes with little distortion and without drawing current from the circuit driving it. Drawing current by any type of detector causes reduction in the *Q* of the circuit and decrease in the selectivity.

In the illustration showing the *Infinite Impedance Detector,* *E* is the modulated carrier voltage and *LC* is a tuned circuit which contributes to the selectivity of the receiver. The plate of the triode is connected directly to the power supply. Condenser C_3 is a virtual short circuit to ground for AC voltages. The load resistance R_1 is in series with the cathode. The condenser C_2 in conjunction with R_1 forms a long RC to the RF frequency but a short RC to signal frequencies.



Infinite-Impedance Detector

Under no signal conditions, the plate current through R_1 provides a bias which keeps the grid very near cutoff. Under signal conditions, when an RF voltage is applied, the positive peak of the modulated voltage changes the condenser C_2 to a higher voltage. This condenser holds its charge between peaks because of the long time constant of R_1 C_2. No grid current flows during the positive peaks because the cathode bias changes at the same time that grid voltage changes and the cathode-grid voltage is never great enough to exceed the cutoff value. When the crest of the modulation peak has passed, the amplitude of the peak decreases and the condenser discharges at the same rate.

The infinite-impedance detector employs a very high input impedance for two reasons.

First, the grid does not draw current. Second, because of the capacity in the plate circuit, the impedance reflected into the grid circuit is purely resistive.

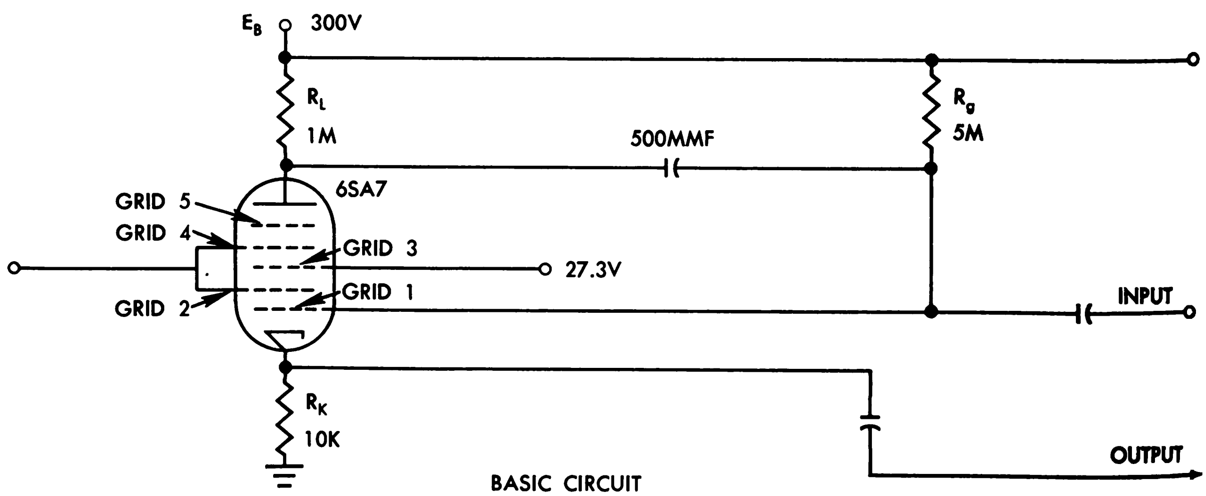
Another characteristic is that negative feedback exists due to the common input and output impedance at signal frequencies. This negative feedback is responsible for reducing the distortion due to characteristics of the vacuum tube.

Distortion is also minimized by the employment of high input voltages. This input voltage can be as large as half the DC voltage on the plate without overdriving the tube. The ability of the tube to handle large signals is due largely to the fact that the bias increases as the signal input increases.

Phantastron Oscillator

The phantastron is used to delay a timing pulse, and is classed as a medium-precision delay circuit. Under power supply voltage changes it is quite stable because its operation depends on the fact that there are specific DC voltage relationships between the tube elements. Any variation of source voltage varies all these voltages in the same proportion, causing a minimum change in the voltage relationship since all voltages are supplied by voltage dividers across the same voltage source.

CIRCUIT. Study the *Phantastron Circuit* illustration. This circuit is usually triggered by a negative pulse at the control grid (grid *1*). Consider first the conditions before the trigger pulse arrives. Grids *2* and *4*, which are tied together inside the tube, are connected to a voltage divider from +300 volts. The normal output of the divider is 86 volts, but with zero control grid voltage and 86 volts on grid *2*, 4 ma of current will flow to grid *2*. This additional current in the voltage divider drops grid *2* to 68.7 volts. Grid *3* is connected to a voltage divider which sets its voltage at 27.3 volts. This grid does not draw current because it is negative with respect to the cathode. The current through the 10K cathode resistor is 4 ma. Since the cathode is 40 volts above ground, the grid *3*-to-cathode voltage is 27.3 − 40 = −12.7 volts. When grid *3* is more than 12 volts negative with respect to the cathode, it prevents electrons from passing it in their journey to the plate. Therefore, since the −12.7 volts is beyond plate current cutoff for grid *3*, no electrons will reach the plate. Grid *1* is connected to E_B through a 5-megohm resistor. A very small grid current flows through the cathode resistor, but the grid voltage may be considered the same as the cathode voltage. In other words, the difference in potential between grid *1* and the cathode voltage is zero. Grid *5*, the sup-



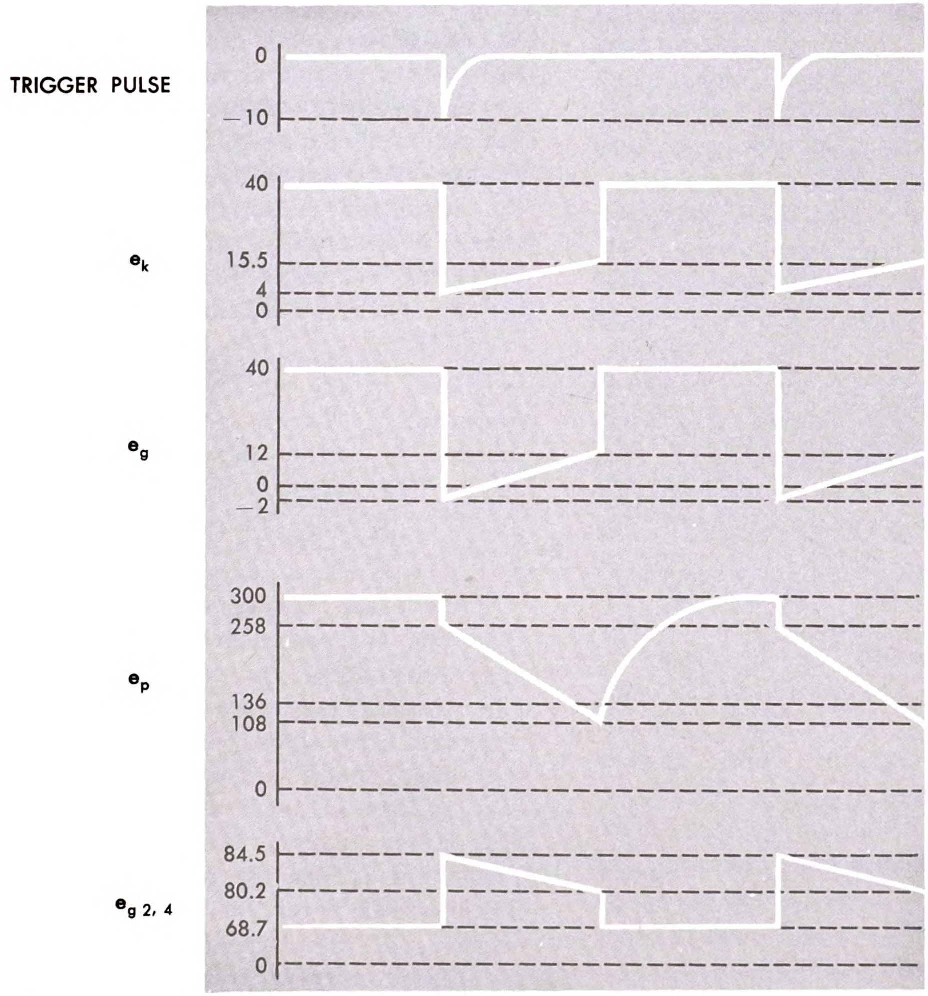
BASIC CIRCUIT

Phantastron Circuit

pressor grid, is connected to the cathode and has no part in the operation of the circuit other than performing the normal function of a suppressor grid.

ACTION. You can best understand the action of the phantastron when it is triggered by dividing it into three successive phases. The first phase is an extremely rapid change which ends at a first balance point. The second phase is a slow linear change which ends at a second balance point. The third phase is the recovery of the circuit to pre-trigger condition.

First Phase. The trigger pulse must have the proper polarity and amplitude to bring grid *3* above cutoff. Since grid *3* is .7 volts beyond cutoff, 1-volt would do it. A negative pulse of 10 volts of amplitude is shown in the *Phantastron Waveshapes* illustration. When this is applied to grid *1*, the grid *1* voltage will drop to 30 volts. Cathode follower action will cause the cathode voltage to drop to 30 volts right with the grid. At this point, examine the relation between the cathode and grid *3*. Their voltage difference is 27.3 —30 or —2.7 volts. Grid *3* is not negative enough to prevent electron flow to the



Phantastron Waveshapes

plate, and plate current will flow as soon as the grid *3* to cathode voltage drops below 12 volts. Regenerative action starts with this plate current flow. The plate current through R_L causes a drop of plate voltage. Grid *1* is capacitively coupled to the plate, so the grid becomes more negative. This drops the cathode voltage still more, which in turn further reduces the voltage difference between grid *3* and cathode. The current to the plate increases, the plate voltage drops, and the grid is driven farther in the negative direction.

The regenerative action just explained is similar to the cutting off of the first multivibrator tube by the second tube. The cutting-off process, however, does not go that far in the phantastron; for, if the current to the plate were stopped, the plate voltage would have to go up. This action would be the opposite to the desired feedback. The increase of current to the plate ends at the first balance point. This balance occurs when the negative-going grid *1* voltage decreases the overall current so much that the plate current cannot continue to increase. Grid *3* is unable to divert enough electrons to the plate to maintain the increase of plate current, so the plate current stops increasing and becomes steady. This occurs at about 2 volts on grid *1*, so the plate has dropped 42 volts before the balance point is reached. Each of these voltages is shown in the curves, under the trigger pulse. The grid voltage has dropped to —2 volts. Since this is a 42-volt change which is caused by the plate, the plate curve shows a 42-volt drop to 258 volts. The cathode voltage drops due to the decrease in plate current until, with —2 volts on the grid, the current is .4 ma, caused by the plate. The grid (*2* and *4*) waveshape shows that E at grids *2* and *4* has gone up because the screen current is decreased, reducing the drop across the voltage divider and raising its voltage 84.5 volts.

Second Phase. Now consider the action resulting from the starting of the second phase. The condenser from plate to grid at this time starts to discharge to the new lower plate voltage. At this time the grid voltage is —2, and the condenser discharges from 300 to 260. In discharging, the grid end of the grid resistor becomes more positive. This increases the plate current, starting another drop in plate voltage. Grid *3* is not involved directly in this stage as it is straight triode action between the plate, grid *1*, and the cathode that affects the circuit operation. The plate voltage drop has an effect opposite to that of the discharging condenser and thereby is degenerative in action. The condenser discharges, the grid goes positive, plate current increases, and plate voltage drops, tending to make the grid go negative. But the plate cannot drop enough to exceed the positive change at the grid because it is the positive grid voltage change that causes the plate voltage drop. So the plate voltage change only counteracts part of the effect of the condenser discharge, slowing the discharge considerably. This negative feedback keeps the condenser discharge in the most linear part of the exponential curve.

The discharge continues at a linear rate as shown by the plate voltage curve in the *Phantastron Waveshapes* illustration. The increase in current raises the cathode voltage to 15.5 volts, and the grid *1* voltage is raised to 12 volts, before the second balance point is reached.

Third Phase. Remember that during all this time, the screen grids (*2* and *4*) have been continually drawing current. As the plate voltage continues to fall, there is a minimum point where 68 volts on the screens will collect more electrons from the space current than a higher voltage on the plate. This occurs at about 108 volts on the plate. This point is also the peak of the current characteristic in the illustration. To the right of this peak, the plate current decreases and the screen current increases. The screen grids take more current than the plate can attract, so the plate current stops increas-

ing and levels off. This leveling off initiates the rapid switchover to recover the circuit to pretrigger condition.

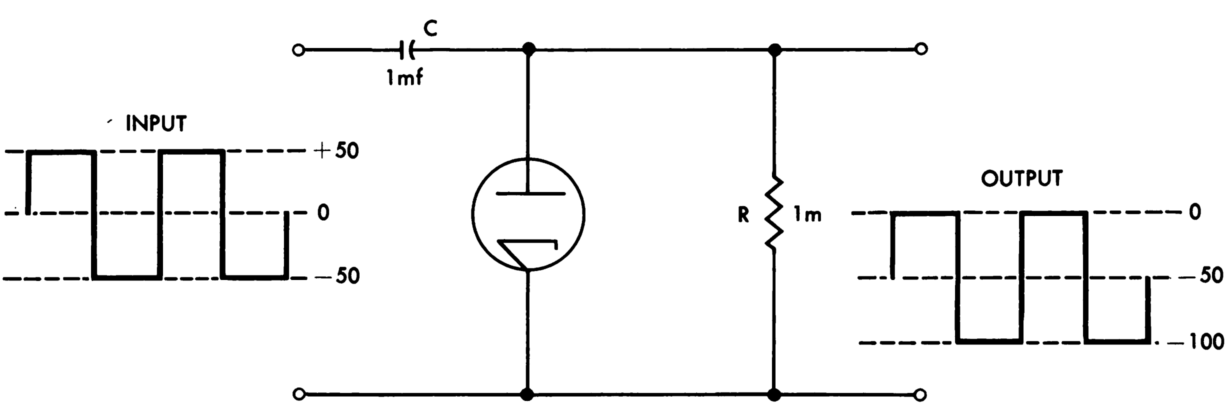
When the plate current levels off, the voltage stops decreasing at the plate. With the counteracting effect of the plate drop removed, the condenser discharge raises the grid *1* voltage at a very rapid rate. The positive-going grid *1* increases the current from the cathode, which raises the cathode voltage. This has the same effect as making grid *3* more negative; so grid *3* reduces the current to the plate. Actually, it is dividing the increasing current in favor of the screen grids. The screen current increases tremendously while the plate current is actually decreased. The plate voltage starts to go in the positive direction. This is coupled to the grid by the 400-mmf condenser to make it go in a positive direction. The cathode voltage is raised some more, which brings grid *3* closer to cutoff, reducing the plate current more and raising the plate voltage. This regenerative action continues until grid *3* is beyond cutoff which stops plate current completely. The third phase of action is then completed; and, to commence a new action, the circuit must again be triggered.

Clamping Diode

The clamping diode is a circuit which shifts a waveshape so that it is all above or all below a certain voltage, which may be zero, any positive voltage, or any negative voltage, depending on the design of the circuit.

NEGATIVE CLAMPER. To analyze a simple clamper in detail, study the elementary *Clamping Diode* circuit illustration. The applied voltage used in the circuit is a square wave with a frequency of 500 c. p. s. and with a variation from plus 50 volts to minus 50 volts in amplitude or 100 volts peak-to-peak. When the positive half-cycle is applied, the condenser charges and the positive voltage causes the diode to conduct. Since the diode has an internal impedance of 1,000 ohms, the 50-volt potential will charge the condenser through a resistance of 1,000 ohms in parallel with 1 megohm, or 999 ohms. The RC time constant will, therefore, be small (approximately .001 second). This small time constant causes the capacitor to charge up fast; so little voltage is felt across the load resistor, resulting in a zero output.

During the negative half-cycle, a negative 50 volts is applied to the circuit, and both the applied voltage and the charge on the condenser cause current to flow in the resistor. During this cycle the diode plate is negative and does not conduct. Since the condenser must discharge through a 1-megohm resistor, the RC time constant will be long (approximately 1 second) and the condenser's discharge will be negligible. In effect, the applied voltage is negative 50 volts and the capacitor is charged to 50 volts. This makes the voltage being felt across the



Clamping Diode

load resistor equal to negative 100 volts; therefore, the output is a negative 100 volts.

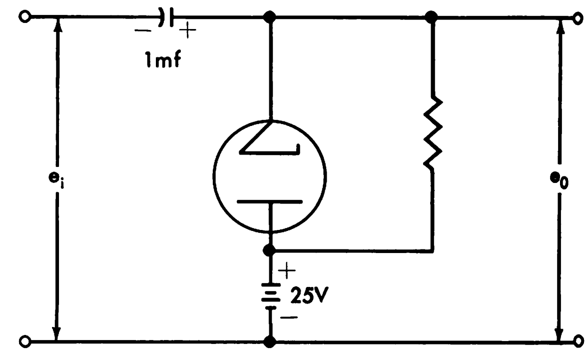
Actually, there is a tiny part of the waveshape left above the zero axis. This bit is the fraction of a volt required to recharge the condenser after each half-cycle. With proper circuit design this can be made negligible.

The clamper diode just discussed may be called a negative clamper because it restores the waveshape in the negative direction. To shift the waveshape in the positive direction, simply reverse the diode connections.

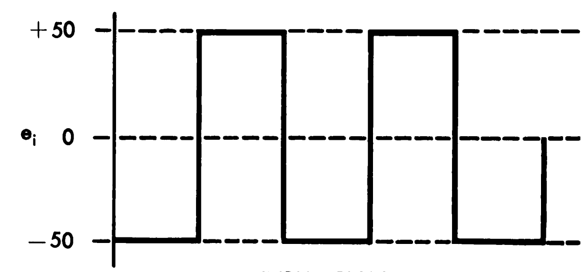
Biased Clampers. A biased clamper shifts a waveshape to an axis other than zero by means of inserting a DC bias of a voltage at which clamping is desired. The only change in the circuit from the clampers just mentioned is the insertion of a DC potential in series with the diode. The circuit of the *Biased Clampers* illustration is a positive clamping circuit that shifts the entire waveshape to the positive side of zero. In operation, the DC source inserts an additional positive 25 volts, which shifts the entire waveshape to 25 volts above zero. Thus, the waveshape which originally varied from minus 50 volts to plus 50 volts, now varies from plus 25 volts to plus 125 volts. Electronically, the shift is due to the DC source charging the condenser to the DC potential, so that when the tube is nonconducting, the condenser has an extra voltage on it in addition to that from the input voltage. For example, in this case, the square wave would charge the condenser to half its peak-to-peak value, or 50 volts. This voltage is sufficient to clamp all the square wave in the output above the zero axis. Since the battery will add a 25-volt charge to the condenser, the total charge will 75 volts. The square wave input varies around this value in the output, or between 75 + 50 and 75 − 50 which is 125 to 25 volts.

Limiters

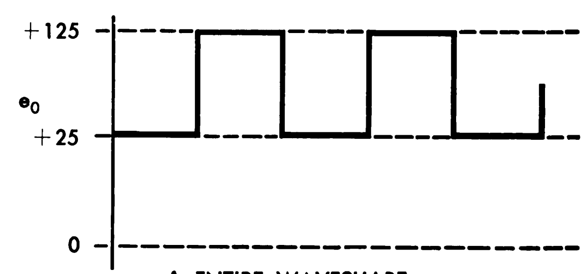
Limiters are circuits which remove either one extremity or the other of an input wave.



POSITIVE RESTORER
WITH POSITIVE BIAS



INPUT SIGNAL

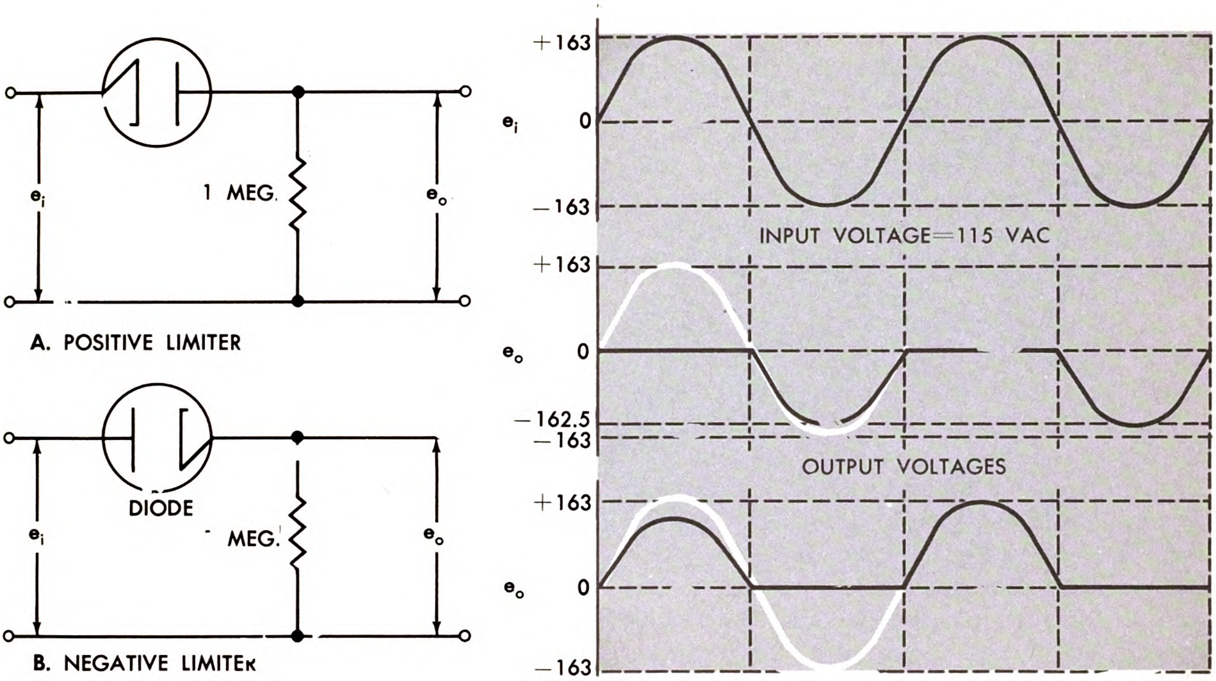


A ENTIRE WAVESHAPE
CLAMPED TO +25 VOLTS

Biased Clampers

Tubes which perform this function are referred to as limiters or clippers.

Limiters are useful in a variety of ways. They are applicable in waveshaping circuits where it is desired to square off the input signal. A sine wave can be converted into a rectangular wave by a limiter circuit. A peaked wave may be applied to a limiter to eliminate either the positive or the negative peaks from the output. Limiters are commonly used to prevent a voltage from swinging too far in either the positive or negative directions.



Series Diode Limiters

SERIES DIODE LIMITERS. Diodes are very useful for limiting since they conduct current only when the plate is positive with respect to the cathode. Diagram **A** in the *Series Diode Limiters* illustration shows a series-connected diode (one in which the load is connected in series with the tube) that is used to limit the positive input cycle of a sine wave. The input voltage is e_i and the output is e_o. When any input voltage is applied to the input terminals of the diode, the output follows the input only when the negative input cycle is applied.

When the positive cycle is applied, the cathode becomes positive with respect to the plate and the tube can not conduct current. Since there is no current flow, there is no voltage developed across the l-megohm output resistor. Therefore, the positive input cycle is limited to zero in the output.

When the negative input cycle is applied, the plate becomes positive with respect to the cathode and, since the plate is at the proper potential, it conducts current. This current flows through the circuit, developing a voltage across the internal resistance of the tube and the load resistor. These two resistances act as a voltage divider and divide the applied voltage. The internal resistance, however, of the diode is small (approximately 1,000 ohms) in comparison to the resistor (1,000,000 ohms). This means that virtually all of the applied voltage appears across the resistor, making the output nearly equal to the positive input. Thus, the diode has limited or clipped off the positive input cycle, and has virtually reproduced the negative input cycle both in shape and magnitude.

When the input terminals are reversed as shown at diagram **B**, the diode limits the negative cycle of the input sine wave because during that cycle the plate is negative with reference to the cathode. During positive cycles, the tube can conduct because its plate is positive with respect to the cathode, and it develops a voltage across the resistor which follows the positive cycle of the input and is equal to it except for the small drop across the tube itself.

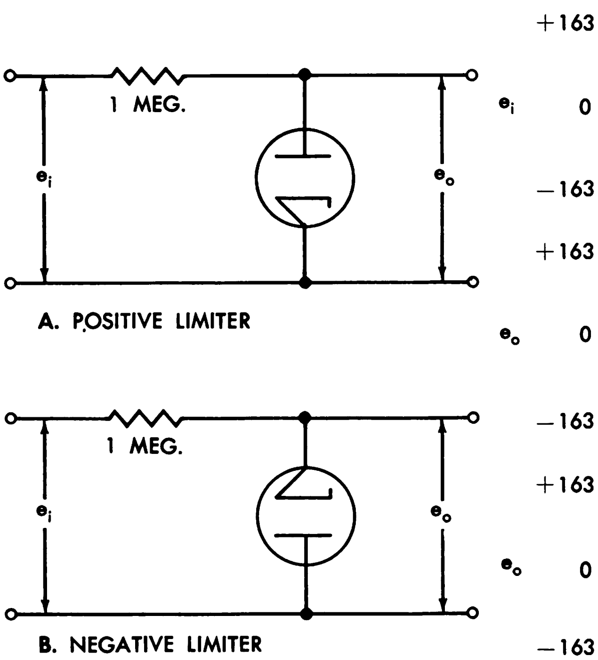
SHUNT DIODE LIMITERS. Another method of using a diode as a limiter is connecting it in parallel with the load. This type of connection is called a shunt diode limiter. In diagram A of the *Shunt Diode Limiters* illustration, the diode is connected to limit the positive input cycle. When the positive cycle of the input voltage is applied to the plate, the diode can conduct. Therefore, as in the case of the series diode limiter, the current flows through a voltage divider consisting of the 1-megohm resistor and the internal plate resistance. This voltage divider divides the applied voltage at a ratio of one thousand to one million. One thousandth of the applied voltage appears across the tube. The remainder, 999 thousandths, appears across the resistor. However, since the output terminals here are across the tube, the output voltage is only one thousandth of the applied voltage. Thus, during the time when the input cycle is positive, the output is limited, or clipped, to practically zero voltage.

During the time the negative cycle of the input voltage is applied to the tube, the plate is negative with respect to the cathode and current does not flow. Therefore there is no voltage drop across the resistor. However, since the tube appears as an open circuit, the applied voltage appears across its output terminals. Thus, the output voltage follows and is equal to the input voltage during the time the negative cycle is applied to the tube.

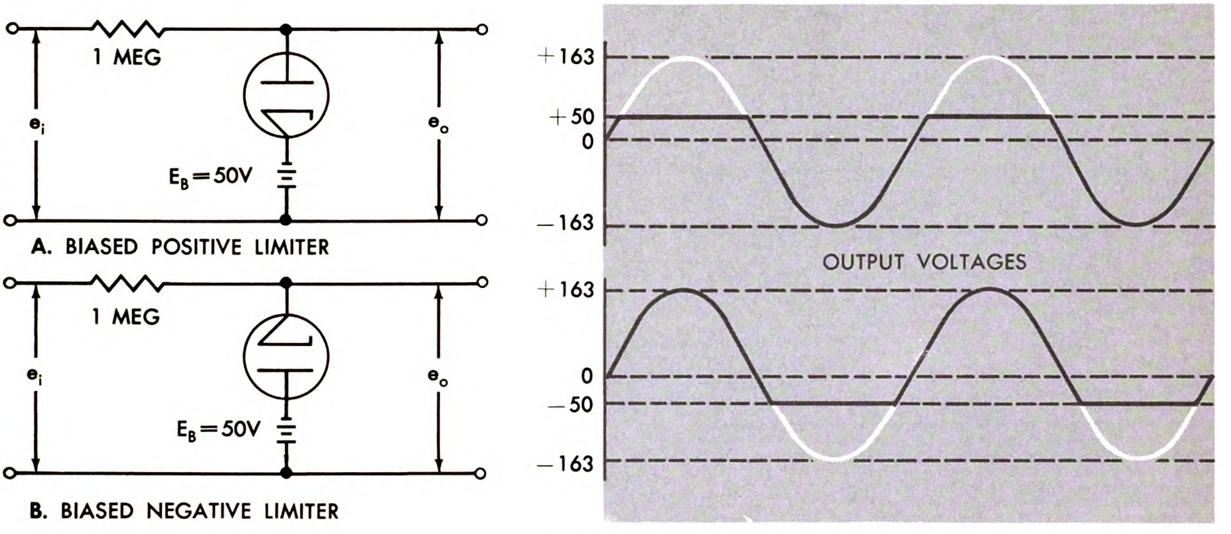
The negative input cycle may be removed by connecting the tube as shown at diagram B. When connected in this manner, the tube conducts when the cathode is negative and virtually all the applied voltage appears across the resistor and a very small amount across the tube, the negative input cycle being reduced or limited practically to zero.

BIASED LIMITING. The input voltage can be limited to some value other than zero by maintaining the plate or cathode at that voltage by means of a battery or a biasing resistor. The two limiting circuits in the illustration, *Limiting to Other Than Zero,* both employ a battery to supply a biasing voltage.

Circuit A is designed to limit the swing of the positive input cycle to +50 volts. The



Shunt Diode Limiters



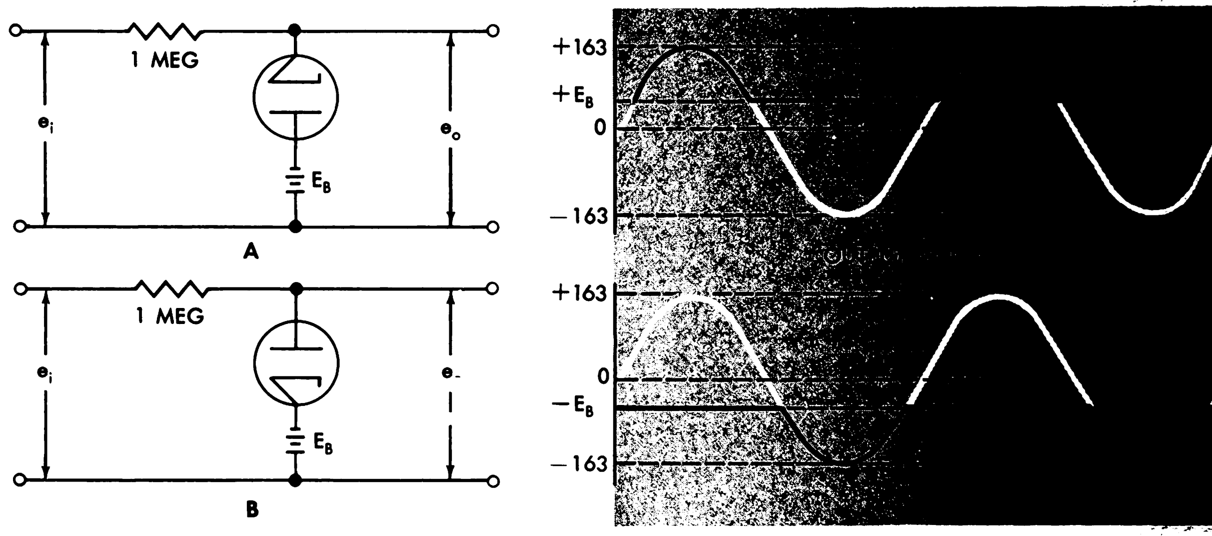
Limiting to Other Than Zero

input voltage e_i is 115 volts effective value and 163 volts peak value. The battery connected in the cathode circuit maintains the cathode 50 volts positive with respect to the plate when the tube is not conducting. As long as the input voltage is less than 50 volts, the tube does not conduct, but just as soon as the input exceeds this amount, current starts to flow and effectively connects the upper output terminal of the battery. Therefore, during the portion of the positive input cycle when the voltage exceeds the 50 volts on the cathode, the output voltage equals 50 volts, and the difference between the input voltage (neglecting the drop across the internal resistance of the tube) appears across the 1-megohm resistor.

In diagram **B**, the battery E_B is so connected in the diode plate circuit as to make the plate 50 volts negative with respect to the cathode. As long as the input is positive or less negative than E_B, the diode acts like an open circuit, and the voltage across the output is equal to the input. When the input becomes more negative than E_B, that is, when the cathode is negative with respect to the plate, the diode conducts current and connects the upper output terminal to the negative terminal of the battery. Therefore, during this part of the input cycle, −50 volts appears in the output, and the rest of the 163 volts input, neglecting the small drop across the tube, appears across the 1-megohm resistor.

Shunt diodes may also be used to limit the amount to which the input voltage can drop. In other words, only the peaks of waveforms are reproduced in the output. In circuit **A** of the illustration *Maintaining Voltage Above Certain Level,* the diode conducts during the entire part of the input waveform that is below the positive voltage of the battery. The output voltage under this condition varies between the positive level of battery voltage and the positive extremity of the input waveform.

In circuit **B** the entire portion of the input waveform above the negative potential of the battery causes the diode to conduct, thus producing an output voltage which varies between the negative level of E_B and the negative extremity of the input. In both cases, circuits **A** and **B**, the difference between the value of E_B and the applied voltage, during the time the diode conducts, is represented by the voltage drop across the 1-megohm resistor.



Maintaining Voltage Above Certain Level

Bootstrap Generator

The bootstrap sweep generator was devised in order to obtain greater linearity and a much larger voltage rise than other generators with the same amount of applied $B+$ supply voltage. The name "bootstrap" refers to the fact that part of the circuit rises above the $B+$ voltage.

So that we can understand the circuit shown in the illustration, *Bootstrap Sweep Generator,* let us first determine which parts are in series and which are in parallel. We can first see that there is a path of current flow from ground through R_K and V_2 to $B+$. We can also see that C is in parallel with V_1.

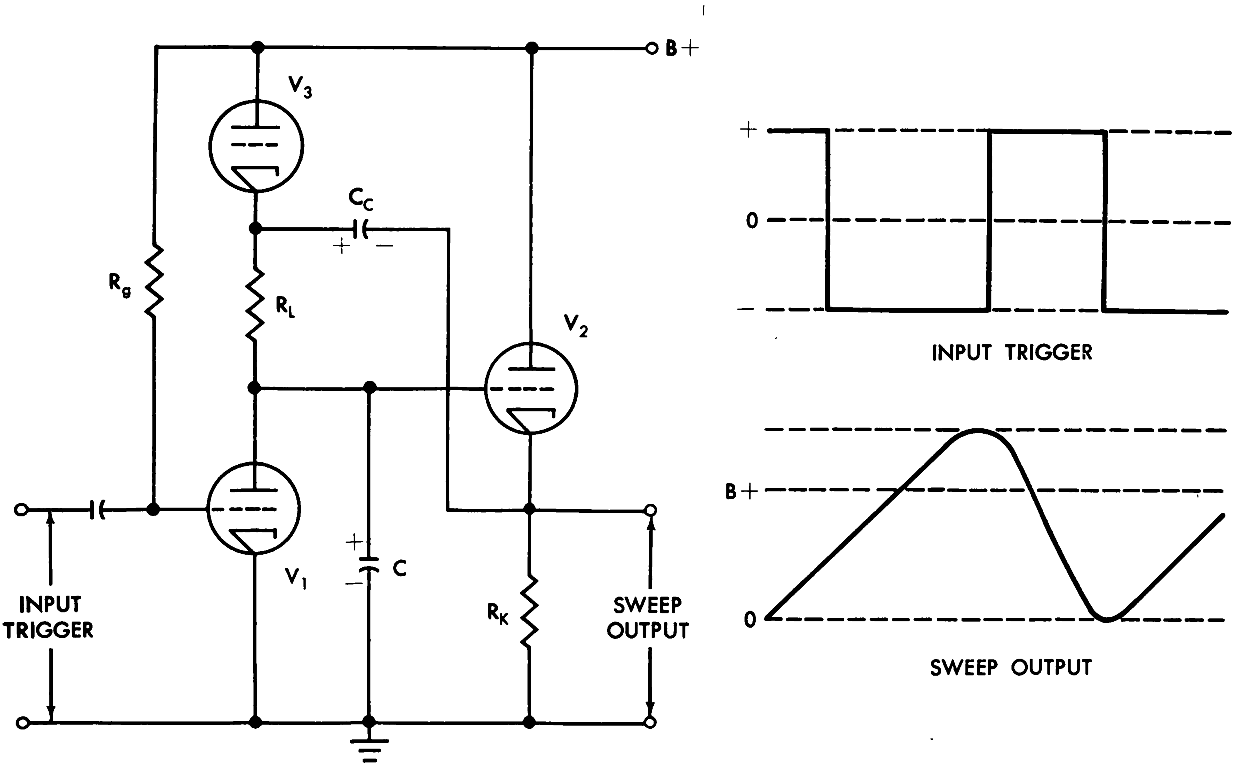
There are two general conditions to this circuit: pretrigger condition, and the condition existing while the negative pulse is being applied to the grid of V_1.

R_g applies a slightly positive bias to the grid of V_1 during pretrigger condition. The input capacitor keeps this DC bias out of the triggering current, but will allow the input trigger to pass. This positive bias on the grid of V_1 will cause it to conduct heavily. Therefore, very little voltage is dropped across the tube, but R_L in series with V_1 has a large voltage drop across it. C in parallel with V_1 has a small voltage drop across it; but V_3, which has little resistance, has virtually none.

V_2 is conducting very little because of the small bias applied from the plate of V_1 and C; hence, R_K, which is in series with V_2, has only a few volts dropped across it. V_2, therefore, has a large voltage drop across it. C_C being in parallel with V_2 will begin charging up to a voltage equal to the voltage drop across V_2.

In the trigger condition, V_1 will cut off because of the negative input trigger. The current flow to $B+$ will then be by way of C, R_L, and V_3. Since C is charging in this condition, current will cease to flow when C has charged to $B+$. The positive voltage, which is on the top plate of C, is felt on the grid of V_2. This makes V_2 conduct harder, increasing the voltage across R_K and decreasing the voltage across the tube. Now notice that R_K, C_C, and V_3 are in a series line from ground to $B+$; therefore, $B+$ will be dropped across all three components.

In the pretrigger condition we found that C_C, being in parallel with V_2, charged [illegible] a high voltage equal to the voltage [illegible] Now, with the increased voltage across [illegible]



Bootstrap Sweep Generator

(hence, a greater current flow), the upper plate of C_C will rise to a voltage above 300. Since this plate is connected directly to the cathode of V_3, the tube will cut off, for its cathode will be positive with respect to the plate.

Since C and R_L are in series with the cathode of V_3 and ground, C will charge toward the voltage on the cathode of V_3. Just as C reaches a point where its charging rate would slack off, the voltage toward which C is charging rises, thus making C maintain a constant linear charging rate.

These linear increases of voltage are applied to the grid of V_2. V_2 increases conduction at a linear rate, which in turn causes the current and voltage across R_K, the output resistor, to increase at a linear rate. Before V_2 is driven to saturation, the negative input trigger ends, causing V_1 to conduct and discharge C, which is in parallel with V_1. This ends the sweep, and we are back in the pretrigger condition. C_C, once charged, holds that charge and needs no discharge path.

RADAR RANGING FUNCTION

The radar ranging function determines the range to the target, tracks the target, and provides data to other sections in the form of signals representing range, range rate, tracking gate, and on-target information.

Delay Lines

In order for target echoes to be displayed at their correct range they are sent through a delay line. This line delays the signal just enough so that the signal does not arrive at the CRT until the time base begins. Since time is required for any voltage change to travel the length of a line, it is possible to

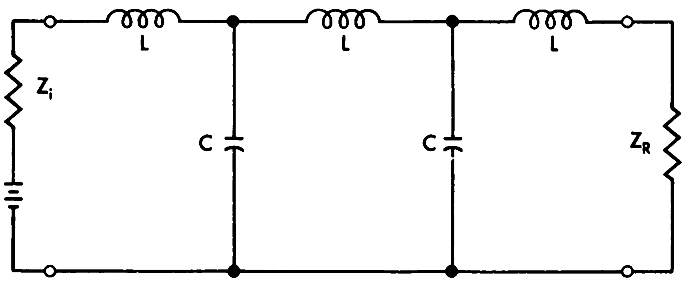
delay the transfer of a voltage change in its travel from one circuit to another.

When DC voltage is applied to a delay line in which the source impedance is matched to the line impedance, half the source voltage appears across the line impedance at the time the battery is connected to the line. This produces a change in voltage across the line. This voltage change travels down the line, charging the line as it travels. When the voltage change reaches the open end, it is reflected back along the line. This is shown in the illustration of an *Open End Delay Line*. All capacitors charge to half the battery voltage when the voltage travels down the line and to full battery voltage when the voltage is coming back. When all the capacitors have been charged to the value of the battery voltage, current stops flowing from the battery.

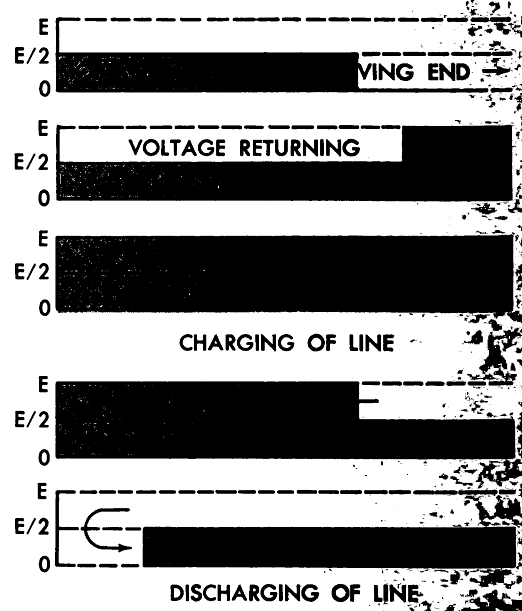
When this line is discharged into a resistance equal to its characteristic impedance (Z_0), the output is a square wave which has a constant amplitude and a duration equal to twice the time for the pulse to travel the length of the line. The amplitude is equal to half the charge voltage because the line discharges through its own impedance and the load impedance (Z_p), which are connected in series. This discharge produces a drop of voltage which immediately travels down the line and back to the starting place as shown in the illustration. As the end of the line is open, the reflection is in phase and discharges the line as it travels back. This means that the duration of the output pulse is twice the time delay (T_d) that can be introduced with an artificial line.

If the load impedance is smaller than the characteristic impedance, most of the energy is dissipated in the line, which results in a lower amplitude. When Z_p is larger than Z_0, the amplitude of the pulse is large. However, several trips of the voltage along the line are required to completely discharge the capacitor. This results in an output of reducing amplitude in the form of steps. The different outputs are shown in the illustration of *Delay Line Outputs*.

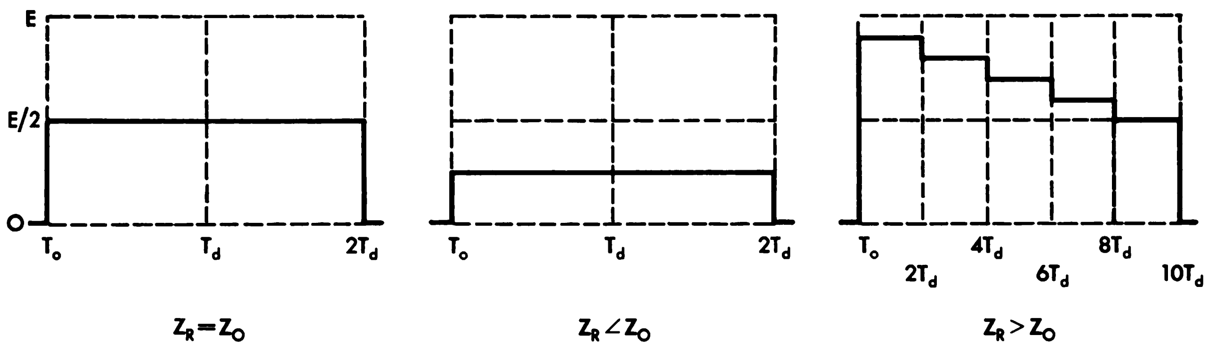
Chapter 4, Section F, contains information on another delay line, the mercury tank sonic delay line.



DELAY LINE



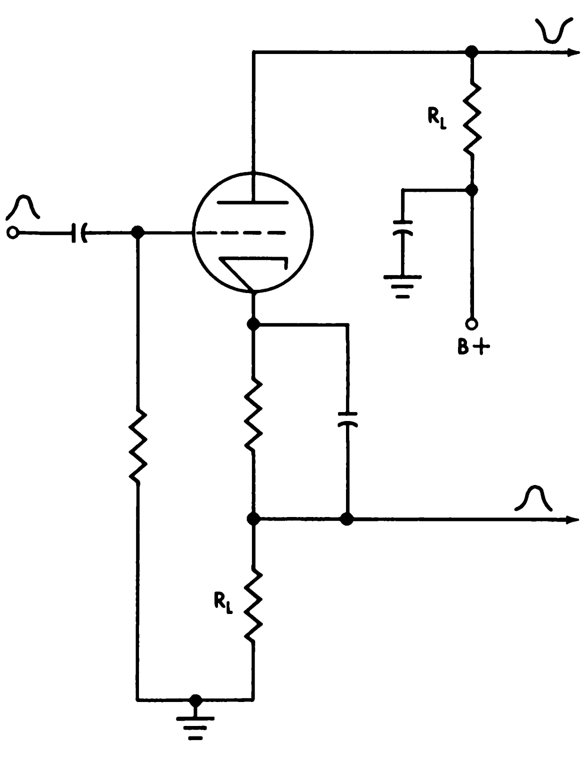
Open End Delay Line



Delay Line Outputs

Paraphase Amplifier

A paraphase amplifier is a combination amplifier and phase inverter, which converts a single input into a push-pull output (two waveshapes of equal amplitude and opposite polarity). There are two types of paraphase amplifiers—the single-tube paraphase amplifier and the two-tube paraphase amplifier.



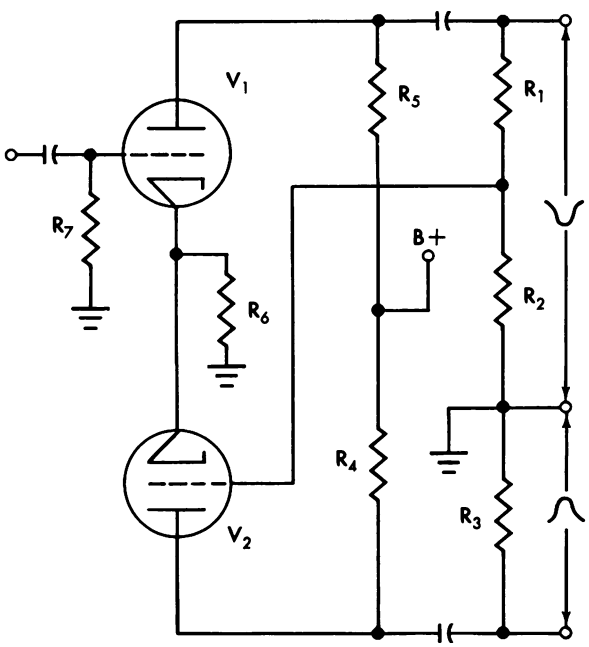
Single Tube Paraphase Amplifier

SINGLE-TUBE. The single-tube paraphase amplifier (see the illustration) output is taken from both the cathode and the plate. The cathode resistor R_L and the plate resistor R_L are the load resistors. These resistors are equal and since the same current flows through both, equal voltages appear across them. The voltages across these resistors are opposite in polarity since the output is taken from the positive end of the cathode load resistor and from the negative end of the plate load resistor.

TWO-TUBE. The two-tube paraphase amplifier consists of one tube which acts as a conventional amplifier, and a second which inverts the output of the first tube. The two tubes in combination then produce two equal output voltages opposite in polarity.

The illustration shows the circuit diagram of a typical *Two-Tube Paraphase Amplifier.* The first tube V_1 amplifies the input waveform shown at its grid and impresses the amplifier output across the voltage divider consisting of R_1 and R_2. The resistor R_2 is of such value that the varying voltage across it has the same amplitude as the voltage on the grid of V_1. This voltage is impressed on the grid of V_2, the phase inverter tube, where it is amplified. Since the plate load resistors, R_5 for V_1, and R_4 for V_2 are equal, the outputs of the two tubes are equal. The phase inverter inverts the phase of the voltage applied to its grid, making it opposite in phase to the voltage output of V_1. Note in this connection that the waveshape in the output of

V_2 is in phase with the grid voltage to V_1. Phase inversion has occurred in V_1 and again in V_2, thus shifting the phase of this voltage back to its original polarity.



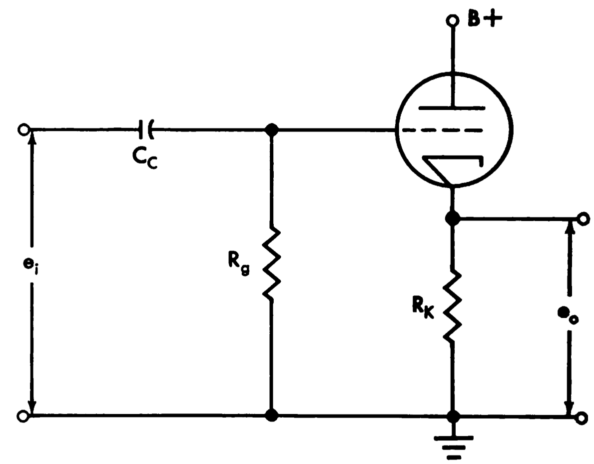
Two-Tube Paraphase Amplifier

Isolation Amplifier

An isolation amplifier is used for the purpose of preventing the impedance of one stage from affecting another stage. This is accomplished by a cathode follower which matches a high-impedance source to a low-impedance device.

Notice in the illustration of a *Cathode Follower* that the signal is applied to the grid of the tube, and that the output is taken across the cathode resistor. Notice also that the absence of a cathode bypass capacitor enables inverse feedback to occur in the circuit.

If a positive signal is applied to the grid, plate current will rise through the cathode resistor, producing a greater voltage drop and making the cathode more positive. Likewise, a negative signal applied to the grid causes a decrease in plate current across the



Cathode Follower

cathode resistor, making the cathode less positive. Thus, the voltage across the cathode resistor "follows" the grid, which, in effect, tends to reduce the voltage difference between grid and cathode as established by the input signal. If no signal is applied to the grid, there is a certain amount of plate current flowing through the cathode resistor. The resulting voltage drop across the resistor establishes the amount of no-signal bias developed. Thus, the signal variation on the grid produces a plate current variation through the cathode resistor which reduces the effectiveness of the input signal. Because of this the voltage gain of a cathode follower is less than one. It does, however, yield a power gain and is therefore classed as a power amplifier.

Coincidence Diode

The coincidence diode, multiplying potentiometers, summing networks, and synchronous vibrators are all discussed in detail under computers, chapter 5.

Integrators

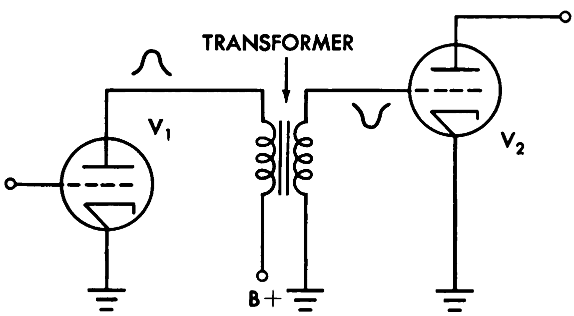
An integrating circuit is one which changes a square wave input to a triangular wave. This can be performed by a RC circuit with a long time constant if the output is

taken across the capacitor. This circuit is discussed in greater detail in chapter 4.

Phase Inverters

A phase inverter is a circuit which produces an output voltage of opposite polarity to the input voltage without distorting the waveshape.

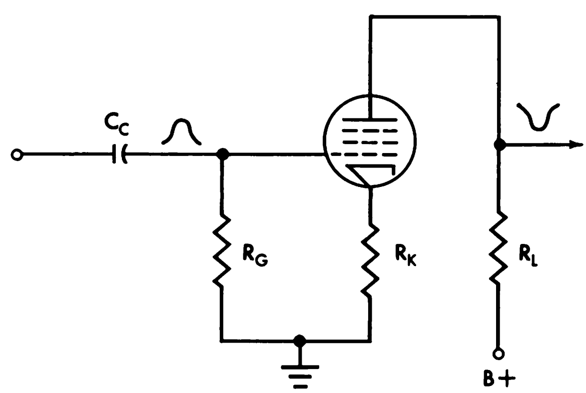
Literally, the term phase inverter is something of a misnomer, since phase is ordinarily associated with time and there is no appreciable time difference or phase shift between the output and input circuits of the ordinary phase inverter. Such a circuit is only an apparent phase inverter; in reality, it is a polarity inverter.



Polarity Inverter

POLARITY INVERSION BY TRANSFORMER. A simple method of inverting the polarity of a waveshape is by a transformer. To understand how a transformer inverts polarity, assume that the pulse at the plate of V_1 in the illustration of a *Polarity Inverter* is positive and that V_2 requires a negative pulse. This requirement can be met by using the polarity-inverting property of an ordinary transformer, since, in all transformers, a current through the primary induces a voltage in the secondary of opposite polarity to the primary voltage. (Of course, if the input or output connections are reversed, the output and input voltages will have the same polarity.) In transformer inversion, it is clearer to think of the output as a voltage whose polarity is inverted relative to the primary voltage, except perhaps when a sine wave signal is used, where the polarity inversion is referred to as a 180° phase shift.

POLARITY INVERSION WITHOUT AMPLIFICATION. Under some conditions, it is necessary to reverse the polarity of a pulse or a waveform without changing its amplitude. Although a transformer may be used, it is much better to use an ordinary RC coupled amplifier with an un-bypassed cathode resistor as shown in the illustration *Polarity Inversion Without Amplification.* This circuit inverts phase since any vacuum tube amplifier with a resistive load has an output of opposite polarity to the input. In other words, a positive-going signal on the grid produces a negative-going signal at the plate. There is no amplification because of the degenerative feedback introduced into the grid by the unbypassed cathode resistor. This degeneration occurs because the cathode voltage rises as the grid voltage rises, preventing the swing of the voltage between the grid and cathode from reaching the amplitude of the applied grid signal.



Polarity Inversion Without Amplification

ANTENNA POSITIONING FUNCTION

The antenna positioning function controls the movement of the antenna during the different modes of operation. There are generally three of these modes—manual, search, and track.

In the manual mode of operation, the radar operator has control over the movement of the antenna.

In the search mode of operation, the antenna automatically scans the area in both azimuth and elevation.

In the track mode, the target has been located and the antenna is automatically and continuously pointed at the target.

Drive Motors

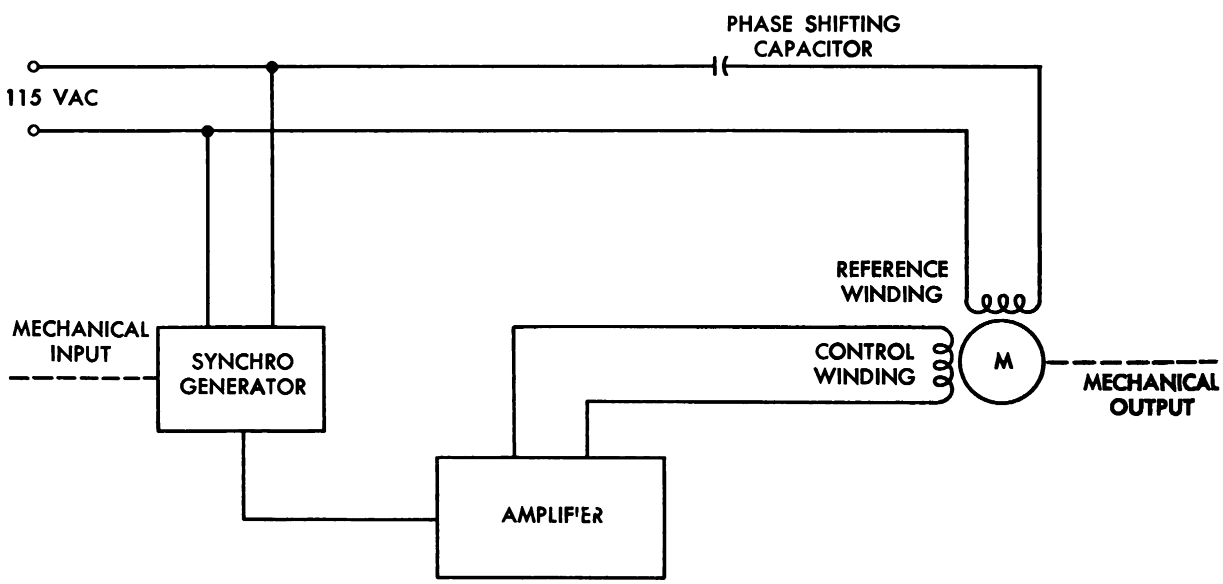
The drive motor of a radar system must be capable of reversing its rotation and changing its speed to minimize any error incurred while tracking a target. This requirement is best fulfilled by the reversible AC motor or induction motor.

In the illustration *AC Induction Motor* the same voltage that is shown being applied to the synchro generator is shifted 90° in phase and applied to the reference winding of the motor (this shift in phase is generally accomplished by a capacitor). If there is no error, the output voltage of the generator will be zero, and the input voltage to the control winding of the motor will be zero. Since under this condition only one phase of voltage is being applied to the motor, no rotating magnetic field exists and the motor remains stationary.

If there is an error, it is an AC voltage which is either in phase or 180° out of phase with the voltage applied to the synchro generator. Thus, the voltage applied to the control winding of the motor will either lead or lag the voltage applied to the reference winding of the motor by 90°. A rotating magnetic field is created and the motor turns in the proper direction to correct the error. As the error decreases the magnitude of the error voltage also decreases; therefore, the signal on the control winding of the motor decreases, slowing the motor down. When zero error is reached, the motor stops. Note that the speed with which the motor rotates at any instant depends upon the magnitude of the error voltage at that instant, and that the speed is continuously decreasing until it becomes zero when the zero error is reached.

Magnetic Clutch

Magnetic clutches are used in most types of induction motors to aid in bringing the



AC Induction Motor

rotor to a quick stop after the energizing current to the starter has been cut off. The brake consists essentially of a brake drum mounted on the shaft extension. The brake drum fits over the brake-coil assembly. Brake shoes are held against the inner side of the drum by spring pressure. The brake-coil assembly is connected in parallel with the field windings. When current is applied to the field, the brake-coil is energized and pulls the brake shoes away from the drum. Rotation of the rotor is then free until the current is cut off. Upon cutoff of the current, the brake-coil assembly becomes de-energized and the spring pressure pushes the brake shoe against the drum, bringing the rotor to a quick stop.

Synchro Resolver

The synchro resolver is used in circuits which solve *vector* problems. In general, three kinds of vector problems can be solved by the resolver. First, it can separate a vector into its two right-angle components—this is called *resolution.* Second, it can combine two right-angle components to produce the vector—this is called *composition.* Third, it can combine many vectors to produce a total vector which is the sum of the given vectors—this is called combination.

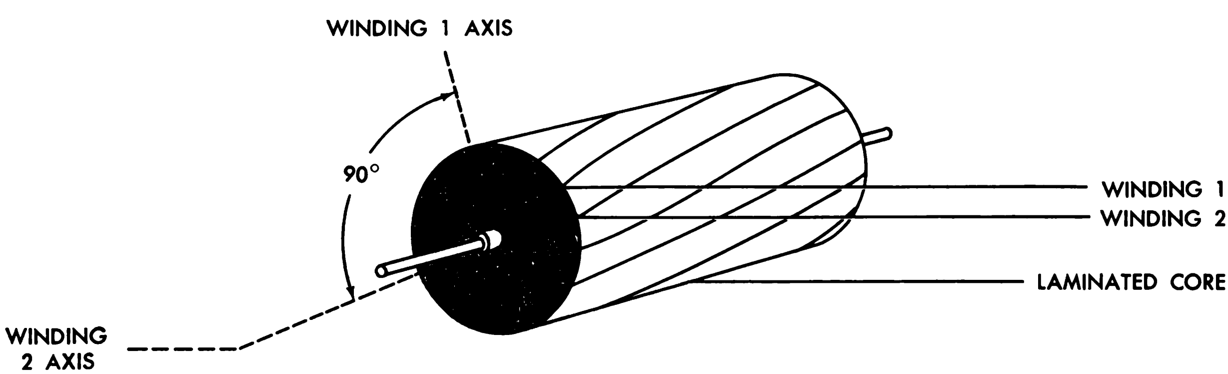
The resolver consists of two major parts—rotor and stator. The rotor core and the stator core are made of special metal laminations called MU-metal. This metal has a very high permeability and a very low hysteresis loss.

ROTOR WINDINGS. Look at the illustration *Rotor Windings.* You can see that the rotor has two separate, independent windings, 90° apart.

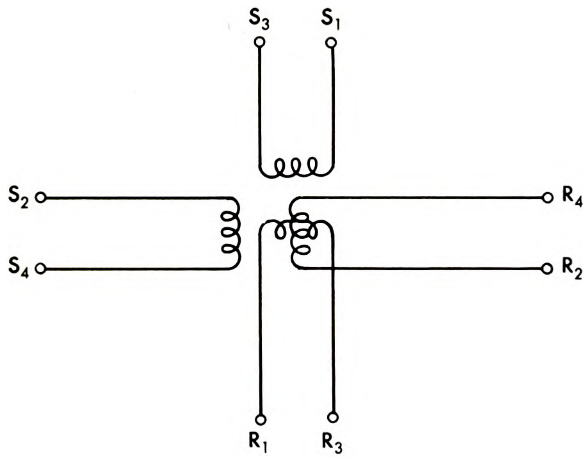
Winding *1* consists of six coils wound parallel to each other and connected in series. Notice that all the coils of this winding have a vertical axis. Winding *2* is likewise made up of six coils. But all the coils of winding *2* have a horizontal axis. One winding has a vertical axis, and the other winding has a horizontal axis. This means the two windings are displaced by 90°, and the 90° displacement prevents any magnetic coupling between the two windings.

STATOR WINDINGS. The stator windings are much like the rotor windings. That is, there are two separate windings, displaced by 90°. Like the rotor windings, there is no magnetic coupling between the two stator windings.

RESOLVER SCHEMATIC. Look at the illustration *Resolver Schematic.* Notice the position of the stator winding—they are drawn 90° apart because that's the way they are wound in the stator. Likewise, the rotor windings



Rotor Windings



Resolver Schematic

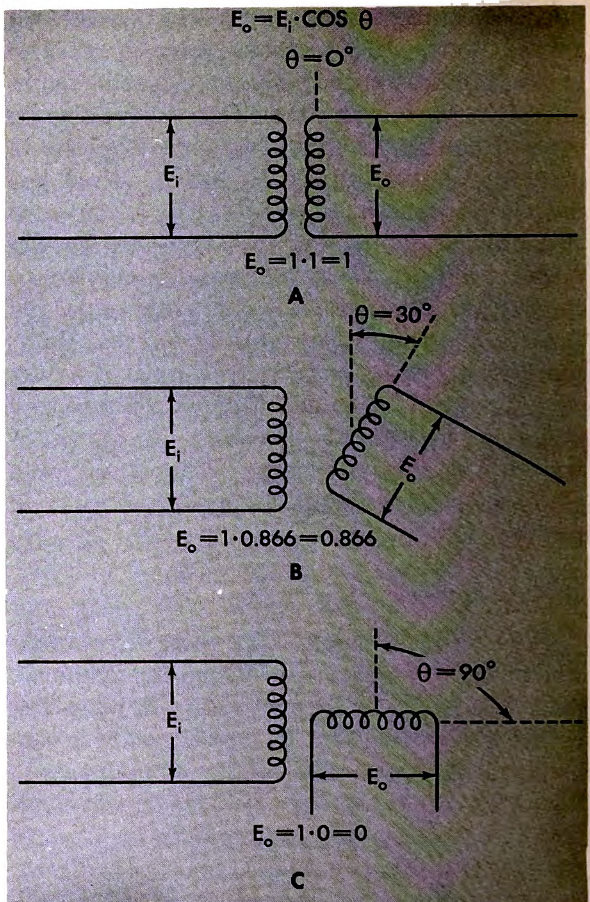
are drawn 90° apart to show their displacement. The schematic in the illustration is usually used in prints. However, many times one of the stator or rotor windings is inactive. When this is the case, the inactive winding is not shown on the print.

RESOLVER PRINCIPLE. The synchro resolver operates on the principle of a transformer. The two stator windings act as primaries, and the two rotor windings act as secondaries. The rotor windings are turned with the rotor shaft. Thus, the secondaries may be at any angle with respect to the primaries.

The illustration, *Resolver Principle, Rotor and Stator,* shows the rotor in three different positions with respect to the stator. This illustration shows only one rotor winding. Similar things happen in both windings—it's easier to explain with just the one winding.

In diagram **A**, the rotor and stator windings are lined up with each other. That is, the angle Θ between rotor and stator is 0°. With the rotor and stator lined up, *all* the flux of the stator cuts the rotor, and the voltage induced in the rotor is maximum.

For example, say the turns ratio between the stator and rotor in diagram **A** is 1:1, and the input voltage E_i is 1 volt. Then the output voltage E_o is also 1 volt, ignoring the small transformer losses.



Resolver Principle, Rotor and Stator

Here's the formula used to find E_o—

$$E_o = E_i \cdot \cos \Theta$$

Then when Θ is 0°, the cosine of Θ equals 1, and

$$E_o = 1 \cdot 1 = 1$$

In diagram **B**, the rotor is turned so that the two windings are 30° apart. This is, angle Θ is equal to 30°. Now, only a *part* of the stator flux cuts the rotor, and the voltage induced in the rotor is

$$E_o = E_i \cdot \cos 30°$$

If the turns ratio is 1:1 and E_i is 1 volt, E_o is

$$E_o = 1 \cdot 0.866 = 0.866 \text{ volt.}$$

In diagram **C**, the rotor is turned so that the two windings are 90° apart. With Θ equal to 90°, the coupling between the stator

and rotor is zero. The voltage induced in the rotor is zero. Or using the formula

$$E_o = E_i \cdot \cos 90°$$

$$E_o = 1 \cdot 0 = 0 \text{ volt}$$

Now, imagine that the rotor in the illustration keeps right on turning. The E_o again rises to a maximum at 180°. And E_o again falls to zero at 270°. Finally, E_o again rises to a maximum at 360°. In short, E_o follows the cosine function for the whole 360°. Also, the E_o reverses phase at 90° and 270°. Thus, the phase of E_o corresponds to the + and − of the cosine function in the four quadrants.

TWO ROTOR WINDINGS. Now see what happens with both rotor windings in place, as in the illustration *Both Rotor Windings.*

Say that winding *1* is the same winding just discussed. Call it the *cosine* winding, because the E_o of winding *1* is the cosine function. However, winding *2* is displaced 90° from winding *1*, and the voltage in winding *2* follows winding *1* by 90°. This means the voltage induced in winding *2* is proportional to the sine function; winding *2* is called the sine winding.

Take diagram **A** as one example. The cosine winding (*1*) is lined up with the stator winding. Angle Θ is 0° and E_{01} is

$$E_{01} = E_i \cdot \cos \Theta$$

$$E_{01} = 1 \cdot 1 = 1 \text{ volt}$$

But with Θ equal to 0°, the sine winding (*2*) is at right angles to the stator winding, and E_{02} is

$$E_{02} = E_i \cdot \sin \Theta$$

$$E_{02} = 1 \cdot 0 = 0 \text{ volt}$$

Use diagram **B** as another example. The whole rotor — both windings — has been turned counterclockwise 30°. The cosine winding has an E_{01} of

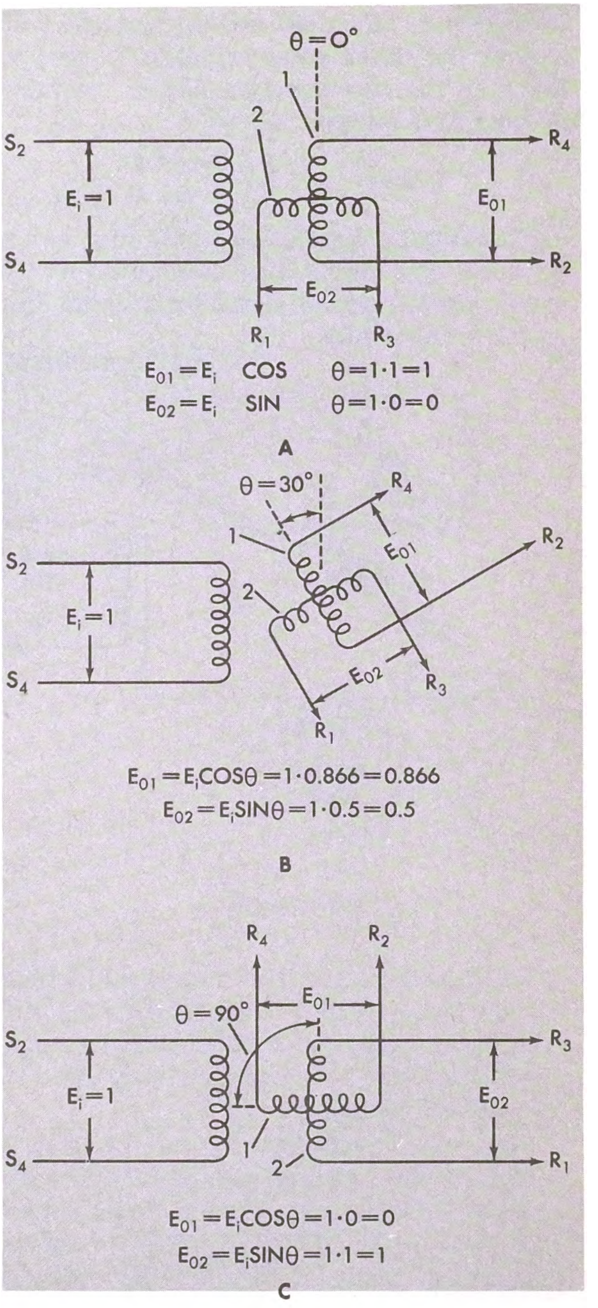
$$E_{01} = E_i \cdot \cos 30°$$

$$E_{01} = 1 \cdot 0.866 = 0.866 \text{ volt}$$

But the sine winding has an E_{02} of

$$E_{02} = E_i \cdot \sin 30°$$

$$E_{02} = 1 \cdot 0.5 = 0.5 \text{ volt}$$



Both Rotor Windings

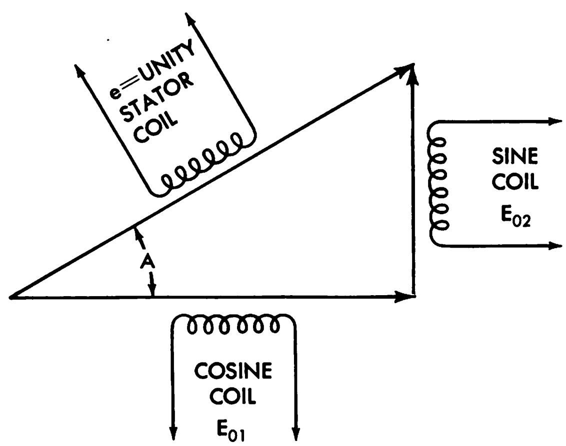
Diagram **C** is a third example. All the equations are shown, and you should be able to get the right answers.

Remember, the important part of the synchro resolver principle is this—the induced voltage in one rotor winding is proportional

to the *cosine* function, and the induced voltage in the other rotor winding is proportional to the *sine* function. When the input voltage E_i is unity—

$$E_{01} = \cos \Theta$$
and
$$E_{01} = \sin \Theta$$

RESOLUTION. Resolution is the first vector problem. The illustration *Resolution, Nr. 1* shows how a resolver separates a vector into its two components.



Resolution Nr. 1

The hypotenuse of the triangle in the illustration is the vector to be solved. The side adjacent to angle A is the horizontal component, and the side opposite to angle A is the vertical component.

First, the stator coil of the resolver is energized with a unity voltage e. This unity voltage represents the vector to be solved. Next, the rotor of the resolver is turned through the angle A. Now the induced voltage in the cosine coil represents the horizontal component because

$$E_{01} = 1 \cdot \cos A$$

The induced voltage in the sine coil represents the vertical component, because

$$E_{02} = 1 \cdot \sin A$$

The two output voltages E_{01} and E_{02} solve the vector problem. Measuring E_{01} gives you the horizontal component, and measuring E_{02} gives you the vertical component.

The illustration, *Resolution, Nr. 2,* shows a more complicated problem in resolution. Three resolvers are connected in a circuit to solve this equation for angle C.

$$\sin A \cdot \sin B - \cos C$$

A unity voltage is fed to the stator coil of the first resolver, R_{e1}, and the rotor of R_{e1} is turned through angle A. This makes the output of the sine coil of R_{e1} equal to the sine of A. The sine of A is fed to the stator R_{e2}. At the same time, the rotor of R_{e2} is turned through angle B. This makes the output of the sine coil of R_{e2} equal to

$$\sin A \cdot \sin B$$

The stator of R_{e3} is also energized by a unity voltage, and the rotor of R_{e3} is turned through the angle C. This makes the output of the cosine coil of R_{e3} equal to the cosine of C. The outputs of R_{e2} and R_{e3} are phased to oppose each other in the heavy lined loop of the illustration. This makes the input to the motor amplifier equal to

$$\sin A \cdot \sin B - \cos C$$

If the input to the amplifier is zero, the equation is solved. Then—

$$\sin A \cdot \sin B - \cos C = 0$$

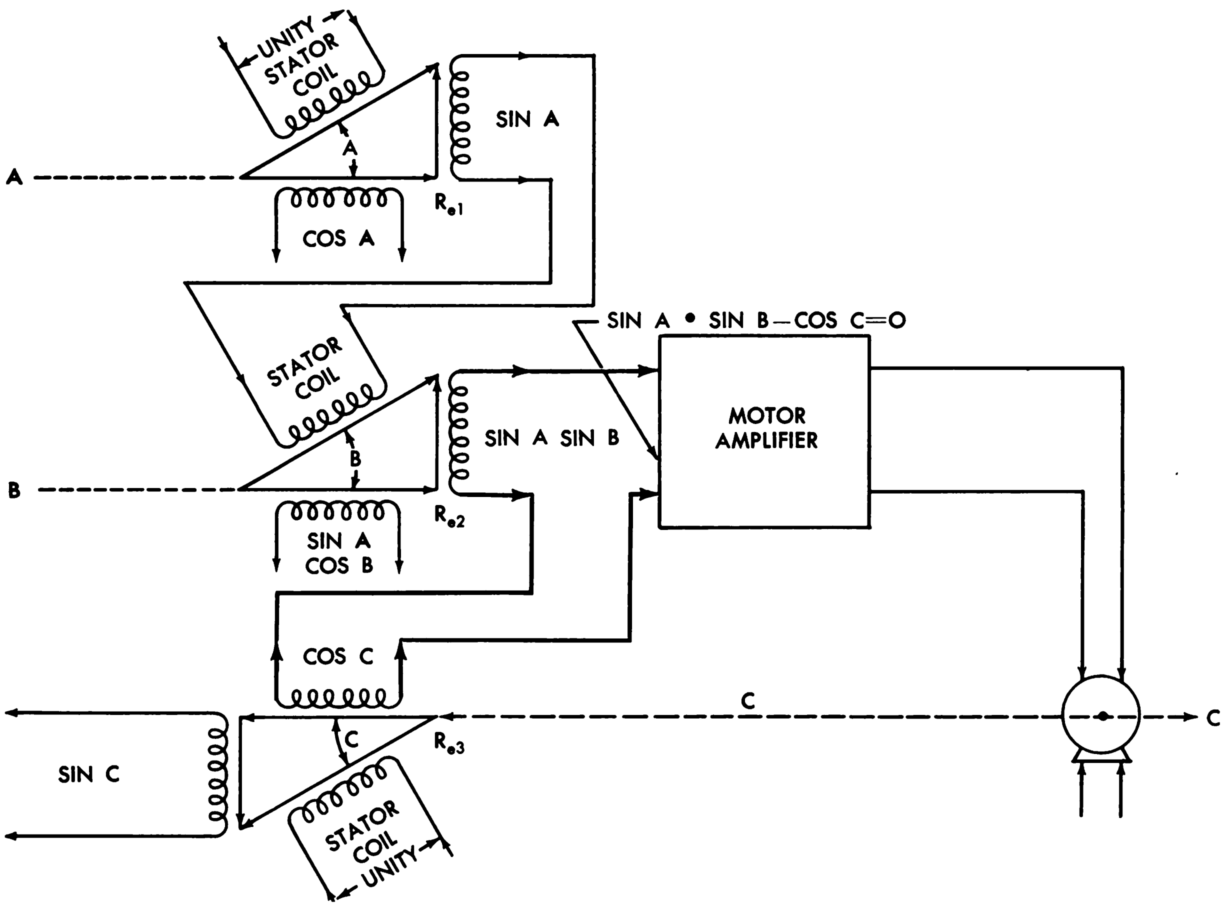
From which, $\sin A \cdot \sin B = \cos C$.

If the input to the amplifier is not zero, the amplifier gets a signal, and the servo drives out a new angle C, and at the same time, repositions the rotor of R_{e3}. The servo drives until the correct angle C is produced. When this correct angle C is produced, the signal is nulled, because

$$\sin A \cdot \sin B = \cos C.$$

Any change in angles A or B immediately produces a servo drive. The servo drives out a new angle C that corresponds to the change in angles A or B. Thus, the system in the illustration continuously solves the equation, $\sin A \cdot \sin B = \cos C$.

COMPOSITION. By composition, you combine the two components of a vector to produce that vector. In other words, the two sides of a right triangle are known. The problem is to solve for the hypotenuse.



Resolution Nr. 2

The resolvers use the Pythagorean theorem,

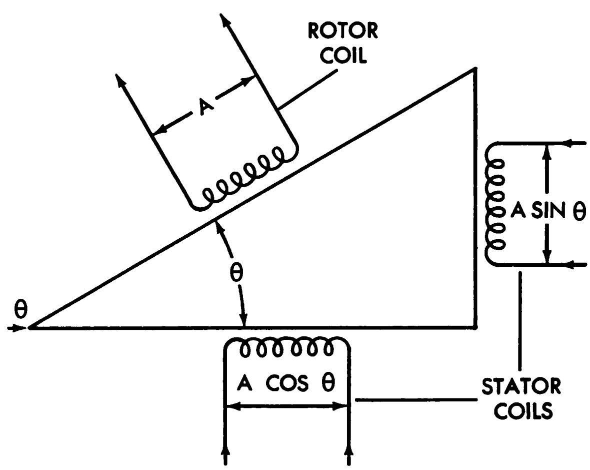
$$a^2+b^2=c^2$$

or,

$$\sqrt{a^2+b^2}=c$$

Now, use the problem in the illustration, *Composition, Nr. 1,* as an example of composition. The two sides of the right triangle are known—*A sine* Θ and *A cos* Θ are voltages representing the two sides. The problem is to solve for the hypotenuse *A*.

The voltages *A sin* Θ and *A cos* Θ are fed to the two stator coils. Each coil produces a flux field proportional to the strength of its voltage. These two flux fields are at right angles because the stator coils are at right angles.



Composition Nr. 1

Right-angle flux fields combine according to this formula—

$$\Phi 1^2 + \Phi 2^2 = \Phi t^2$$

or,

$$\sqrt{\Phi 1^2 + \Phi 2^2} = \Phi t$$

Notice that the resultant flux Φt is proportional to the square root of the sum of the squares of the two individual fields. Since the two fields are produced by the voltages $A \sin \Theta$ and $A \cos \Theta$, the resultant flux is proportional to

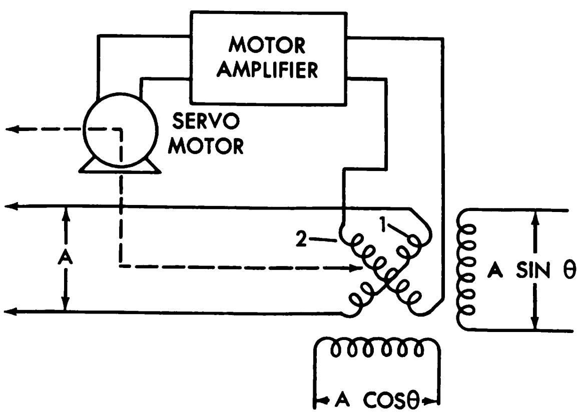
$$\sqrt{(A \sin \Theta)^2 + (A \cos \Theta)^2}$$

The resolver rotor is turned through the angle Θ so that the axis of the rotor coil comes into alinement with the axis of the resultant stator flux. The voltage induced in the rotor increases as the coupling increases and at correspondence the rotor voltage is proportional to the vector A where

$$A = \sqrt{(A \sin \Theta)^2 + (A \cos \Theta)^2}$$

The problem is solved; voltage A is the hypotenuse of the right triangle.

Now to see how the resolver continuously solves for A, refer to the illustration *Composition, Nr. 2.*



Composition Nr. 2

The two voltages representing the two sides of the right triangle are again applied to the stator coils of the resolver. The resultant flux induces a voltage in the rotor coil. This rotor voltage is proportional to the hypotenuse A.

But there is a second rotor coil now in use, coil *2*, which supplies a signal to the motor amplifier to drive the servo.

When coil *1* is cut by all the resultant flux, the output of coil *1* is proportional to A, but coil *2* is at right angles to the flux. Result—no voltage is induced in coil *1*, and no signal goes to the amplifier. The servo does not drive because the circuit has solved the problem.

When coil *1* is not cut by all the flux, coil *2* feeds a signal to the amplifier; and the servo drives out angle Θ and drives the resolver in response. The servo continues to drive until coil *2* is at right angles to the flux and the signal is zero. When the signal is zero, coil *1* cuts all the resultant flux and its voltage represents A. Also, when the signal is nulled, the servo shaft position represents angle Θ.

If either input voltage changes, the resultant flux changes, and coil *2* feeds a signal to the amplifier. The servo drives until the signal is again nulled. At this point, the induced voltage in coil *1* represents the new value of A. Thus, the servo system continuously solves for A and Θ.

COMBINATION. By combination you combine two or more quantities to solve certain trigonometric equations. Actually, combination is a process of evolution and composition, both taking place simultaneously.

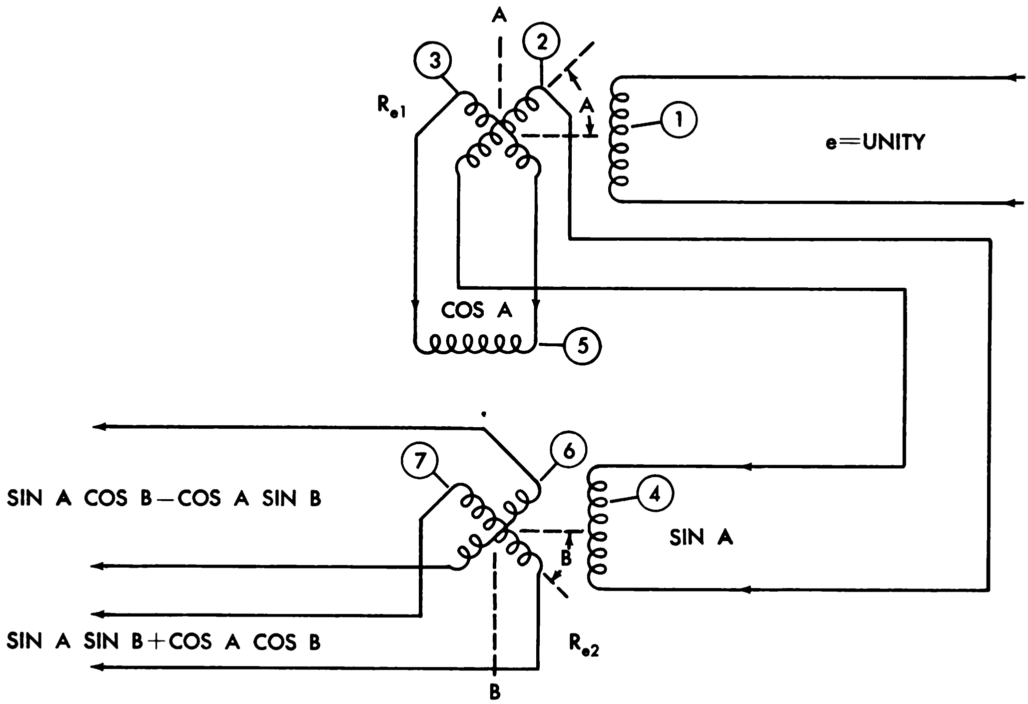
Here's an example that will show you how combination works. Say that you have two angles A and B, and you want to produce these functions of the angles—

$$\sin A \cos B - \cos A \sin B$$

and,

$$\sin A \sin B + \cos A \cos B$$

The illustration, *Combination* [illegible] shows the circuit that produces these two functions. The position of the [illegible] represented by angle A and the [illegible] angle B.



Combination Circuit

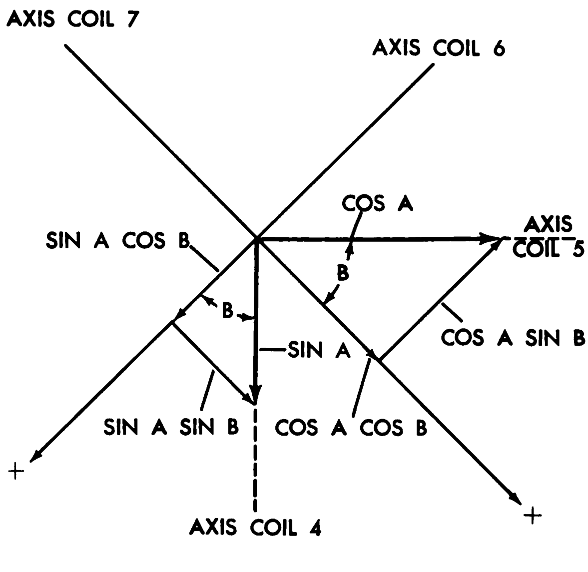
Coil *1* in the stator of R_{e1} is energized by a unity voltage. Result—coil *2* in the rotor produces the sine of A. Coil *3*, also in the rotor, produces the cosine of A.

The two outputs of R_{e1}, *sin A* and *cos A*, are fed to the two stator coils of R_{e2}. Result—coil *4* sets up a flux proportional to *sin A*, and coil *5* sets up a flux proportional to *cos A*. The resultant flux in R_{e2} cuts the R_{e2} rotor coils *6* and *7*. The induced voltages in coils *6* and *7* depend on the rotor position, angle B.

To see exactly what happens in R_{e2}, look at the vectors in the illustration, *Combination Vectors*. The two heavy vectors are the flux fields set up by coils *4* and *5*. The vertical flux vector represents *sin A*, and the horizontal flux vector represents *cos A*.

Next, locate the axis of the R_{e2} rotor coils. Coil *6* is displaced from the *sin A* vector by angle B, and coil *7* is displaced from the *cos A* vector by angle B.

Now see what voltages are induced in coils *6* and *7*.



Combination Vectors

First, analyze coil *6*. Coil *6* has two flux components that are parallel to its axis, $\sin A \cos B$ and $\cos A \sin B$. These two flux components cut coil *6* and induce the coil-*6* voltage. Therefore, the coil *6* voltage is proportional to

$$\sin A \cos B - \cos A \sin B$$

That is the first function of angles A and B. Now analyze coil *7*. Coil *7* also has two flux components that are parallel to its axis, $\cos A \cos B$ and $\sin A \sin B$. These two flux components cut coil *7* and induce a voltage proportional to

$$\cos A \cos B + \sin A \sin B$$

That is the second function of angles A and B.

Thus, the two rotor outputs of R_{c2} are proportional to the two functions of angles A and B.

Rate Gyro

Rate gyros are covered in detail in Chapter 4, Sections C and D.

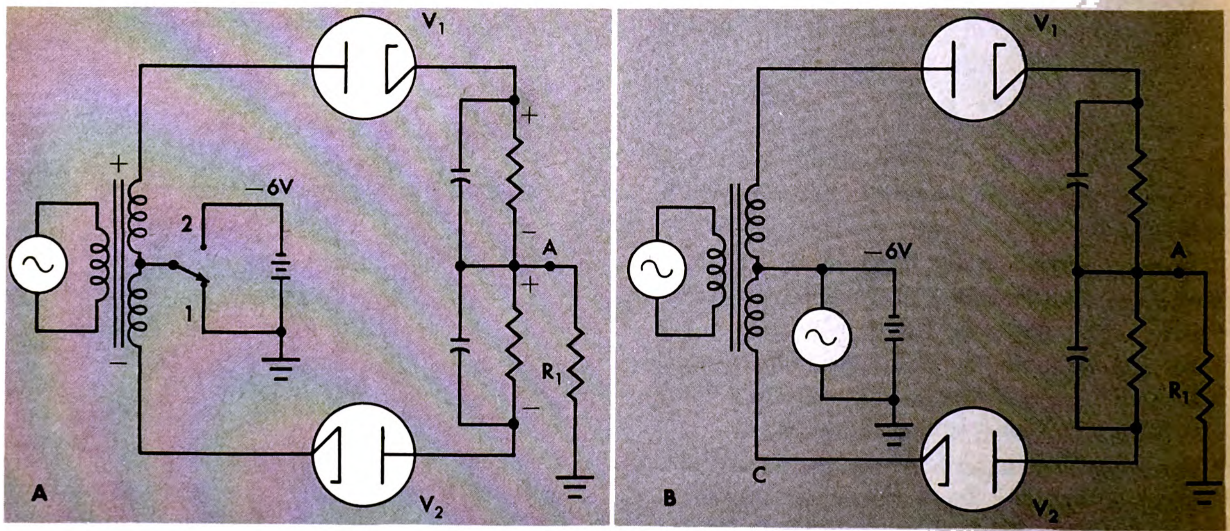
Phase Detector

Phase detection is the comparison of an AC signal of unknown phase to a signal of known phase. The illustration, *Phase Detector* shows a simplified schematic of a phase detector.

You can see from diagram **A** that, if an AC voltage is applied to the plate of V_1 and the cathode of V_2 with the instantaneous polarity as indicated, both tubes will conduct equally. The currents will cancel in the common resistor (R_1), and point A will be at ground potential. Although the tubes both conduct simultaneously, they conduct during only one half of the AC cycle applied.

If *S1* is switched to position *2*, a bias voltage will be applied to the center tap of the transformer. The currents in the common resistor R_1 will not cancel, and point A will be at a potential equal to the bias applied to the center tap.

Consider the AC signal applied to the primary of the transformer to be the reference phase, and apply an AC signal of unknown phase to the center tap in addition to the minus 6-volt bias. You can tell whether the unknown AC signal is in phase or out of phase with the reference voltage by determining whether the voltage across R_1 is above or below the minus 6-volt level. Why this is so can be seen from diagram **B**. If the unknown AC signal is in phase with the reference voltage and is in phase with point B (going positive as B goes positive), then



Phase Detector

V_1 will cause more current to flow through R_1 than V_2. This causes the potential at point *A* to decrease in a negative direction. If the unknown AC signal is in phase with point *C*, V_2 conducts more and the voltage across R_1 increases in a negative direction. You can, therefore, determine whether the applied AC signal is in or out of phase with the reference voltage by determining whether point *A* is above or below the bias level.

Magnetic Amplifier

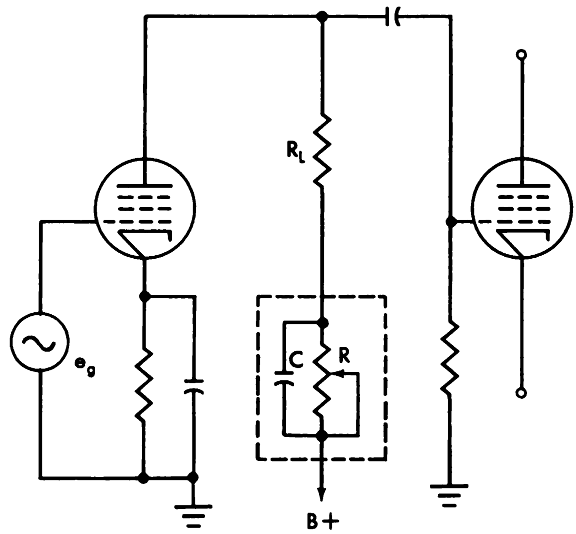
Refer to Chapter 4, Section E.

Frequency Compensation Networks

Frequency compensation is necessary to counter the effects of attenuation and phase shifting which will occur when an amplifier has to pass a broad band of frequencies.

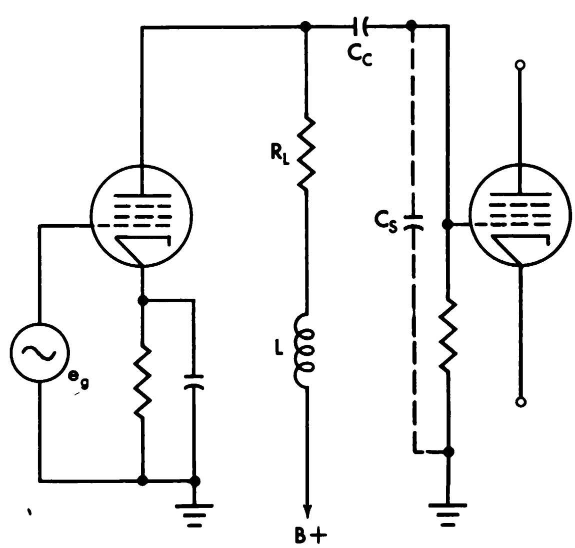
LOW FREQUENCY COMPENSATION. When low frequency distortions occur, it is necessary to increase the output for low frequencies and provide an opposite phase shift without affecting high frequencies. This is accomplished in the plate circuit of the amplifier where the plate load is varied for low frequencies only.

A satisfactory means for improving low frequency amplification is to use a low frequency compensating circuit similar to the one shown in the illustration, *Low Frequency Compensation Circuit.* This low frequency compensating circuit improves response by using a parallel RC circuit in series with the load resistor. Since *C* is comparatively large, it offers practically no reactance to middle and high frequencies and, therefore, does not affect these frequencies. At low frequencies, however, its reactance is high. This reactance in parallel with *R* produces an impedance which, when added to R_L, increases the total load impedance. Since the compensating circuit produces a larger load impedance at low frequencies, the gain of the stage is higher at these frequencies.



Low Frequency Compensation Circuit

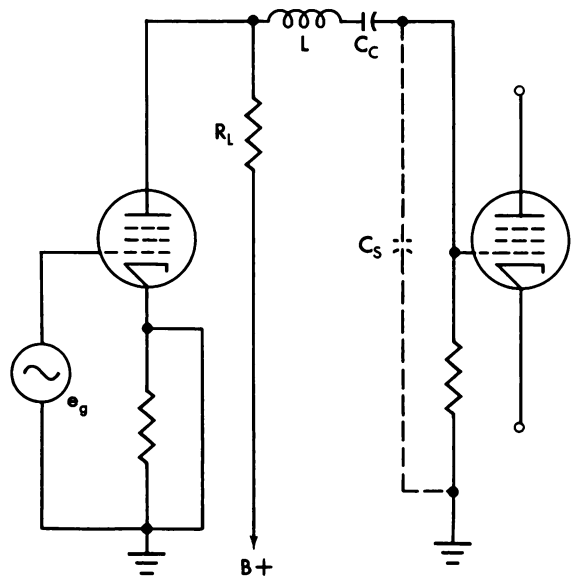
HIGH FREQUENCY COMPENSATION. The use of small load resistors in plate circuits of amplifiers does not completely solve the problem of improving high frequency response. Therefore, carefully designed high frequency compensating circuits are often used. In the illustration, *Shunt Peaking Coil,* a low reactance coil (*L*) is placed in series with the load resistor (R_L) and in shunt with C_s. At low and middle frequencies the reactance of



Shunt Peaking Coil

the coil is low. However, at high frequencies, the reactance is high and the gain of the amplifier increases. This increase in gain offsets the loss caused by the shunting effect of C_s. The value of the coil is usually chosen so that it forms a resonant circuit with the shunt capacitances at a frequency where the gain is low and increases the gain at this frequency. In this way, the shunt peaking coil extends the band of frequencies which the video amplifier can amplify uniformly.

The series peaking coil is another type of high frequency compensating circuit. As shown in the illustration, *Series Peaking Coil*, a coil (L) is connected in series with



Series Peaking Coil

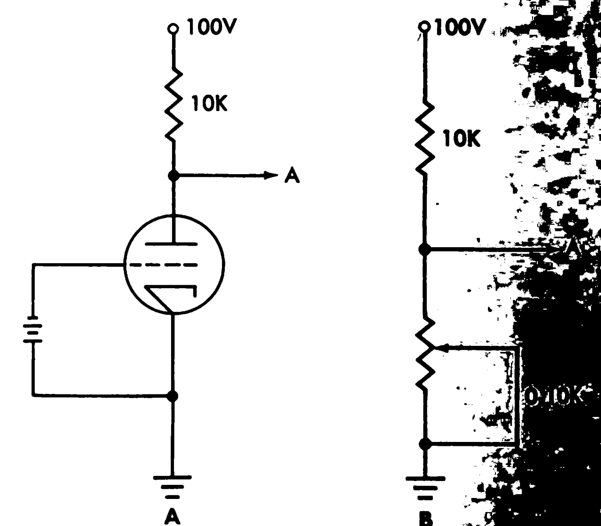
the coupling capacitor C_c. At high frequencies, the inductor L, capacitor C_c, and capacitor C_s form a series resonant circuit. The increased current in this resonant circuit causes an increased voltage drop across C_s, and a greater high frequency voltage is applied to the grid of the following stage.

The two methods of high frequency compensation are often combined in video amplifiers to increase the bandwidth at the high frequency end.

Variable Resistance Triodes

A triode tube, like most other circuit components, has DC resistance. This DC resistance may be defined as the opposition the tube offers to the flow of current at any instant. The DC plate resistance will vary with the bias, but the AC plate resistance [illegible] remains fairly constant.

Because the DC resistance of a triode will vary as its grid voltage is varied, a triode in series with a resistor can be considered as a voltage divider, with the tube acting as a variable resistor. Compare the two circuits shown in the illustration, *Triode as a Variable Resistor*, and note the similarity.



Triode as a Variable Resistor

In diagram B, if the 0-10K rheo[illegible] shorted out, the voltage at A would [illegible] If the variable arm of the rheostat [illegible] midpoint, putting 5K of resistance [illegible] cuit, the voltage at A would be 33[illegible] the entire 10K rheostat were [illegible] voltage at A would be 50 volts [illegible] voltage at A is the applied vo[illegible] minus the drop across the fixed [illegible] As more resistance is added [illegible] the current flow decreases, the [illegible] the fixed resistor decreases, and [illegible] at point A increases.

In diagram A then, if the DC resistance of the tube is increased, the current flow through the circuit will decrease, the voltage drop across plate load resistor will decrease, and the voltage at *A* (since it equals the applied voltage minus the drop across the plate load resistor) will increase. The voltage at point *A* is the voltage drop across the tube and is equal to the plate current times the DC resistance of the tube. If the bias is increased (more negative) the DC resistance of the tube will increase and the voltage at *A* will increase (more positive).

This brings out a very important fact. As the grid voltage moves in a negative direction (more bias), the plate voltage moves in a positive direction; conversely, if the grid voltage moves in a positive direction (less bias), the plate voltage moves in a negative direction, or becomes less positive.

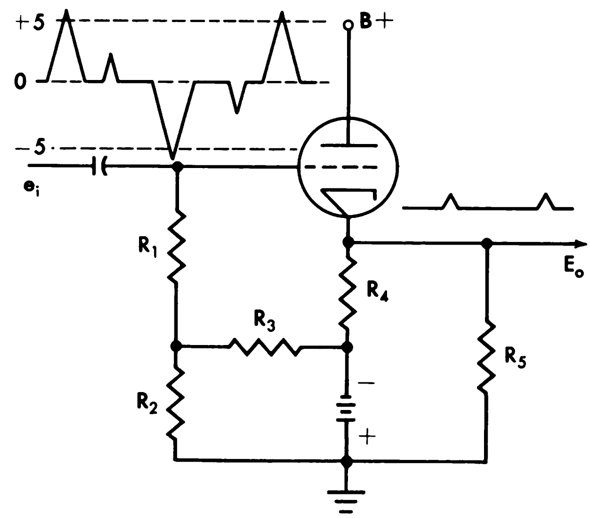
MISSILE SERVO FUNCTION

The Falcon is a small supersonic homing air-to-air guided missile. Because it is a homing type missile, certain information must be presented to it prior to launching. The missile servo function in preparing the missile for launching during an attack presents the necessary information.

Peak Detector

It is often necessary to use peaked pulses separated by a certain time interval to maintain proper synchronization of a system. To accomplish this a peak detector is generally employed.

The illustration, *Peak Detector,* is a simplified schematic presentation. Study it carefully and you will note that its operating characteristics are those of a limiter and cathode follower combined. The battery in the cathode circuit furnishes the bias to the tube through the two voltage divider networks of R_1, R_2, and R_3, and R_4 and R_5. By choosing the proper value of these resistors, the grid can be biased at a negative potential in respect to the cathode.



Peak Detector

Assuming that the cathode is biased at minus 5 volts and the grid is biased at minus 10 volts, we have made the grid 5 volts more negative than the cathode. Therefore, the tube will be cut off. To allow the tube to conduct, a signal with an amplitude of 5 volts or more must be applied to the grid. When a positive pulse of this amplitude is applied, the tube will conduct. Since the tube is connected like a cathode follower, the cathode will follow the peak of the pulse, giving us an output equal to the peak of the applied pulse and in phase with it.

Low-Pass Filter

A low-pass filter passes all currents of frequencies below its cutoff frequency to the desired circuits, and opposes or diverts the flow of all currents of frequencies above the cutoff point.

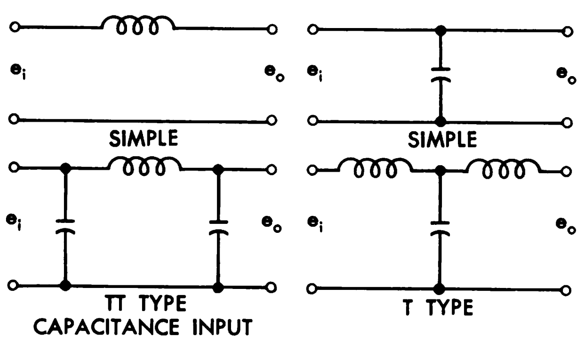
The simplest type of low-pass filter is either an inductor in series with the load or a capacitor in parallel with the load. The inductor in series with the load will attenuate the current as the frequency increases, and the capacitor in parallel with the load diverts the current from the load as the frequency increases. Neither of these produce

very sharp attenuation, and in order to obtain better attenuation both the inductor and capacitor are used. To further improve the sharpness of attenuation, additional capacitors and inductors are used. The illustration, *Low-Pass Filters,* shows how they can be combined.

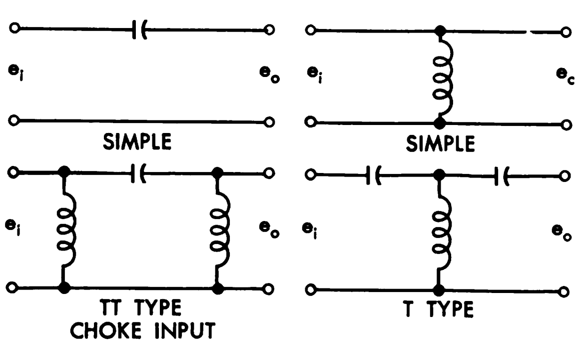
High-Pass Filter

A high-pass filter passes all currents of frequencies above its cutoff frequency, and cuts off or diverts all currents of frequencies below this point.

The simplest type of high-pass filter is a capacitor in series with the load, which opposes the low frequency currents, or an inductor in parallel with the load, which diverts the low frequency currents from the load. Like the simple low-pass filters, neither of these will offer very sharp attenuation. So again we combine the two as shown in the illustration, *High-Pass Filters.*



Low-Pass Filters



High-Pass Filters

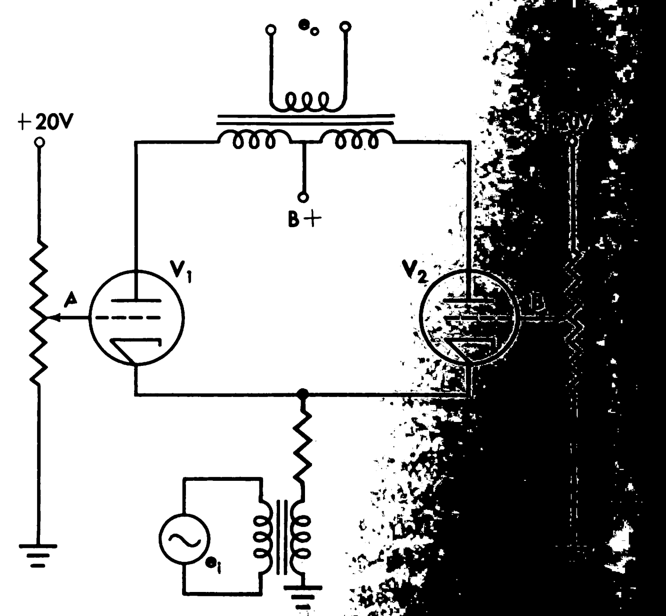
400-CPS Modulator

The modulator receives a DC input (usually from a phase detector) and puts out a 400-cycle AC voltage.

Study the illustration of the *Basic Modulator Circuit.* Now assume that grid *B* is at ground potential and grid *A* is at plus 20 volts. Under these conditions, if a negative signal is applied to the common cathode, V_1 will conduct and V_2 will cut off. At the same time, the plate of V_1 will go negative and will be in phase with the applied signal. Similarly, if grid *A* is grounded and grid *B* is at plus 20 volts, a negative signal applied to the common cathode will cause V_2 to conduct and V_1 to cut off. The important thing to observe is that the plates are connected by a push-pull output transformer. If a current flows from the plate of V_1 to the center tap, the output phase will be reversed.

ARMAMENT CONTROL

The armament control enables the pilot to select the type of armament and amount of armament to be used during an attack run.



Basic Modulator Circuit

CHAPTER 8

PRINCIPLES OF INFRARED

The interest in using infrared radiation techniques in military search, detection, tracking, and similar equipment is increasing all the time. Because of this increase of interest, it is probable that personnel in the armament career field will find themselves more closely associated with infrared waves in the near future.

INFRARED RADIATION

Infrared radiation is just another type of radiant energy transmitted through space in the form of electromagnetic waves. These infrared (IR) waves, like RF waves, have a definite wavelength at a specific frequency. The wavelength is the distance from a peak of one portion of the wave to a corresponding peak on the next like portion of the wave.

Spectrum

The *rainbow* of color produced by sunlight passing through a crack into a darkened room is a visible color spectrum. This phenomenon is the result of refraction and dispersion of light rays, and it exemplifies the basic principle of spectrography.

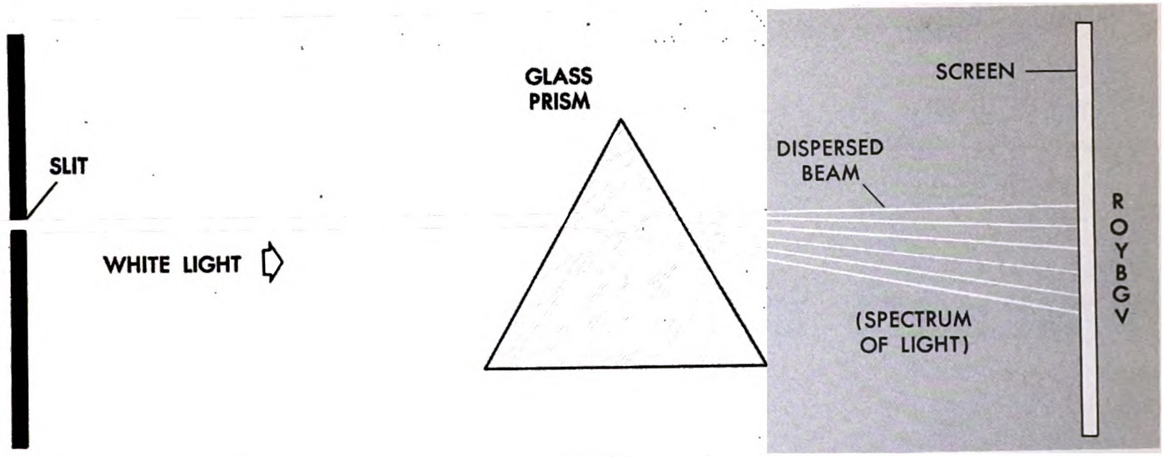
Actually, white light is composed of light rays of many hues blended together. When such light is passed through a narrow opening or slit into some diffracting medium, rays of each different color are diffracted at different angles and spread out into six principal bands of color (refer to the illustration *Dispersion of Light Through a Prism*). These colors red, orange, yellow, green, blue, and violet compose what is known as the visible light area of the spectrum.

The color of light is determined by its frequency of vibration with red having the lowest frequency and violet having the highest frequency. This can be seen in diagram **B** of the *Spectrum Chart* illustration. Diagram **A**, however, shows us that the spectrum extends below and above the visible light portion, and that the frequencies of electromagnetic radiation extend from near zero upward.

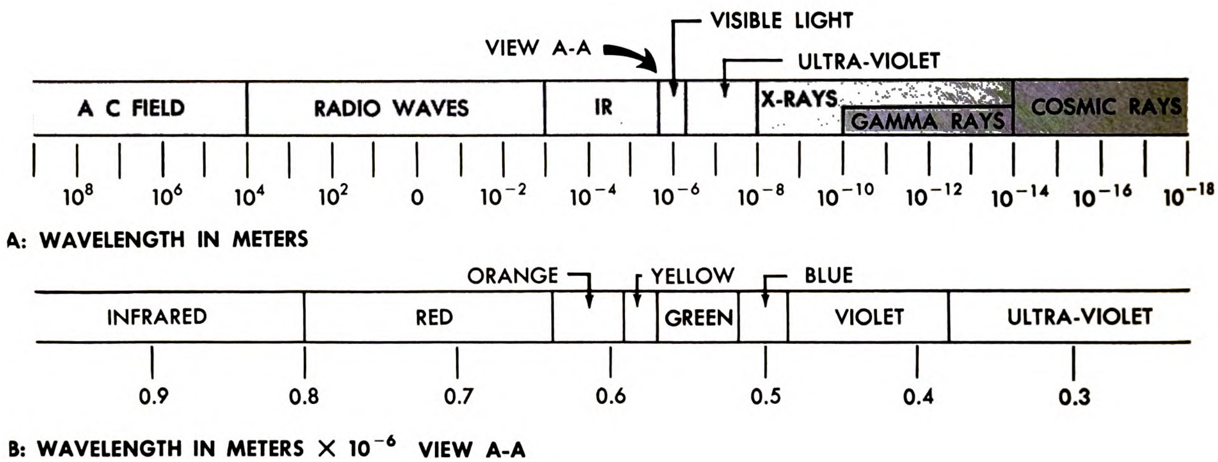
The main differences between the various types of electromagnetic radiation are the methods of generation and detection. Infrared radiation is the name given the region of the electromagnetic spectrum from 300×10^{-6} to 0.7×10^{-6} meters. The low frequency of the IR energy spectrum extends toward the high frequency end of the radar region, and the high frequency end of the IR region extends to the visible light region.

Generation and Measurement

The best source of IR energy is from thermal radiators or hot objects, and is generally called "radiant heat." Examples of



Dispersion of Light Through a Prism



Spectrum Chart

thermal radiators are the sun, light bulbs, and burning wood.

All objects are continuously radiating and absorbing energy from surrounding objects. When radiant energy strikes a body, the energy of a specific wavelength is either passed through, absorbed, or reflected by the object. If a body is a good absorber of energy of a specific wavelength, it will also be a good radiator of the same wavelength but a poor reflector. A body that is a poor absorber and radiator of energy of a specific wavelength will be a good reflector of this wavelength. Some objects will be good absorbers and radiators for a group of wavelengths. Examples of this are the earth, foliage, water, and ice. Some good reflectors over part of the IR band are highly polished metal surfaces and metal flake paints such as aluminum paint.

It is more difficult to find a substance that transmits only a portion of the IR band; however, some crystalline materials such as potassium bromide, potassium chloride, and sodium chloride pressed into a solid form may be polished and used for this purpose. In the near portion of the IR band (portion nearest visible red light) quartz, magnesium

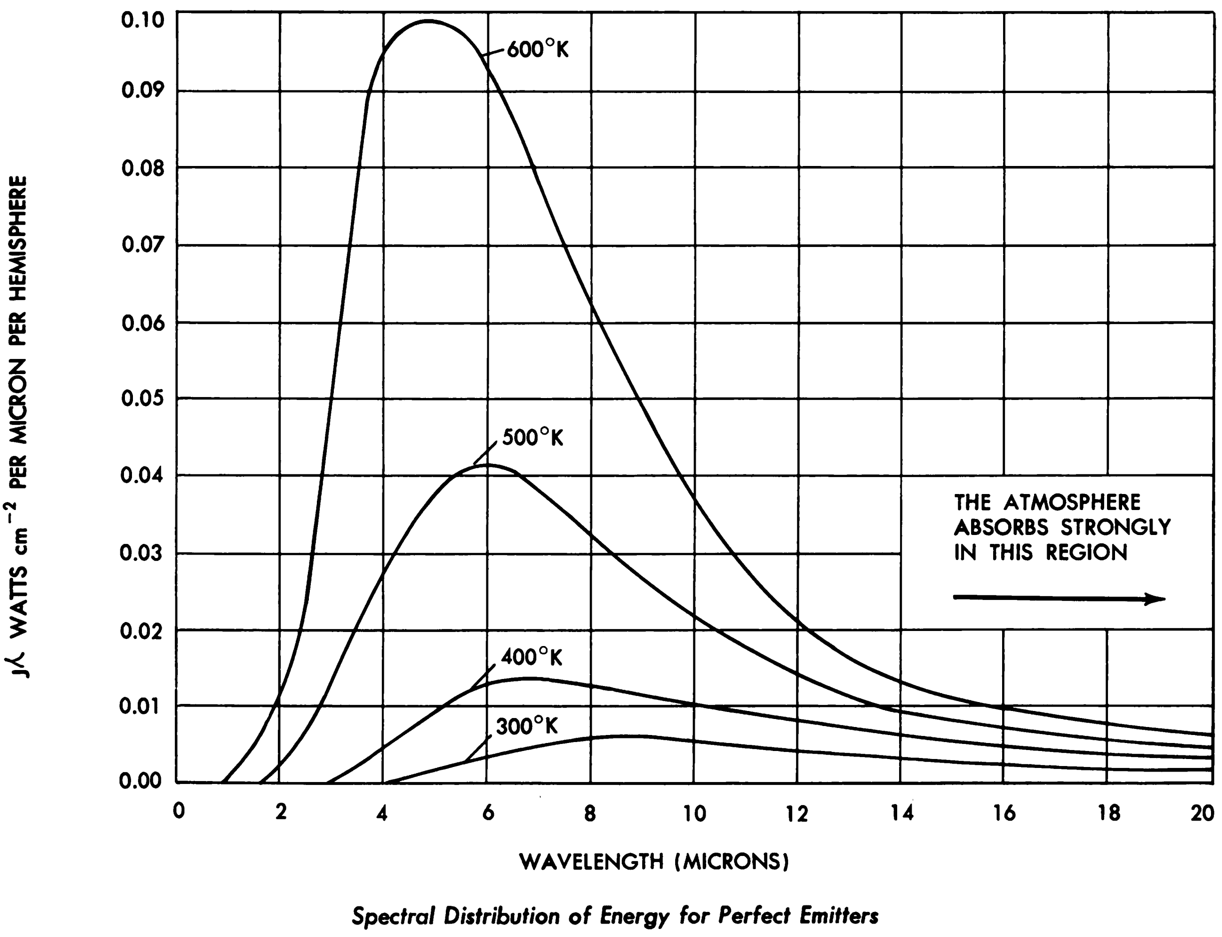
oxide, arsenic trisulfide glass, and a few other substances may be used to pass IR energy. Generally these materials are transparent to only a very narrow portion of the spectrum.

BLACK BODY. The theoretical reference body to which all bodies are referenced in measuring their absorption and radiation characteristics is called a "black body." Kirchhoff's law defines a perfect black body as an object that absorbs all the electromagnetic radiation falling on it, neither transmitting nor reflecting any of it, irrespective of whether or not the radiation is in the visible portion of the spectrum. The absorption and emission of such a black body is unity. The spectral distribution of black bodies is shown in the illustration *Spectral Distribution of Energy for Perfect Emitters.*

We can now use a perfect black body as defined by Kirchhoff's law as a standard of measurement in discussing how infrared energy is radiated, measured, and detected. Perfect black bodies rarely occur, although most materials except metals and a few transparent substances may be considered as black bodies over most of the IR band.

A close approximation to a black body can be realized through the use of an enclosure of uniform internal temperature with a very small opening in one wall. A large cave with a narrow mouth and uniform internal temperature is an example of a close approximation to a black body.

EMISSIVE POWER. The emissive power of a source is of interest so we can have some idea of how much energy a typical aircraft



Spectral Distribution of Energy for Perfect Emitters

might radiate. The total emissive power of a surface of unit area is the amount of energy of all wavelengths radiated per second into a solid angle of 2π steradians (a hemisphere). Stefan-Boltzmann law states that the amount of energy radiated by a body is dependent upon the temperature of the radiating body. Stated mathematically:

$$E = \varepsilon \, \delta \, T^4$$

where, E = total emissive power. (Watts cm^{-2} per hemisphere)

ε = total emissivity of surface. (One is for a perfect black body and is an empirically determined amount of one or less for all physical bodies. The ε for any body is the ratio of its emissive power at temperature T to the emissive power of a black body at the same temperature.)

δ = Stefan-Boltzmann constant. (5.7 times $.16^{-12}$ watts $cm^{-2}deg^{-4}$)

T = absolute temperature in degrees Kelvin.

The above equation describes the total emissive power of a unit area of a surface into a solid angle of 2π steradians. The intensity of radiation (I) of a surface is another term that needs to be defined. It is the emissive power per steradian normal to the surface and is expressed

$$I = \frac{E}{\pi} = \frac{\epsilon \, \delta \, T^4}{\pi}$$

From this equation, it can be seen that increasing the temperature of a body by a factor of two will result in increasing the intensity of radiation by a factor of sixteen.

The illustration of *Spectral Distribution of Energy for Perfect Emitters* shows that, as the temperature of the radiating body increases, the peak of the spectrum shifts toward the shorter wavelengths. This phenomenon is explained by Planck's Distribution Law and, in a simplified form, by the Rayleigh-Jeans equation which is accurate within one percent for $\lambda T > 77$ cm degrees.

The Planck Distribution Law describes the emission of thermal radiation from a unit area of a black body into a hemisphere as follows:

$$J\lambda = \frac{C_1}{\lambda 5} \cdot \frac{1}{e^{c2}/\lambda T - 1}$$

where, $J\lambda$ = emissive power of a unit area in the wavelength interval

C_1 = 3.732 times 10^{-12} watt cm^2.

C_2 = 1.436 cm-degree

λ = wavelength in centimeters.

T = absolute temperature of the radiating body. (degrees Kelvin)

Rayleigh-Jeans equation is:

$$J\lambda = \frac{C_1}{C_2} \cdot \frac{T}{\lambda} \cdot 4$$

Taking the derivative of Planck's Distribution Law with respect to wavelength and setting it equal to zero and then solving for wavelength maximum, Wien's Displacement Law is obtained. This law is valid for small values of wavelength and temperature only. Wien's Law is:

$$\lambda \, m.T = 2892$$

where, $\lambda \, m$ = wavelength (in microns) for maximum radiated energy

T = absolute temperature (degrees Kelvin).

The intensity of radiation from a source at some distance away from the source is of interest since we need to know how the intensity of IR radiation will vary with range. In general the variation of intensity of radiation with range is given by the inverse square law:

$$\frac{I_1}{I_2} = \frac{R_2^2}{R_1^2}$$

where, I = intensity of radiation.

R = range

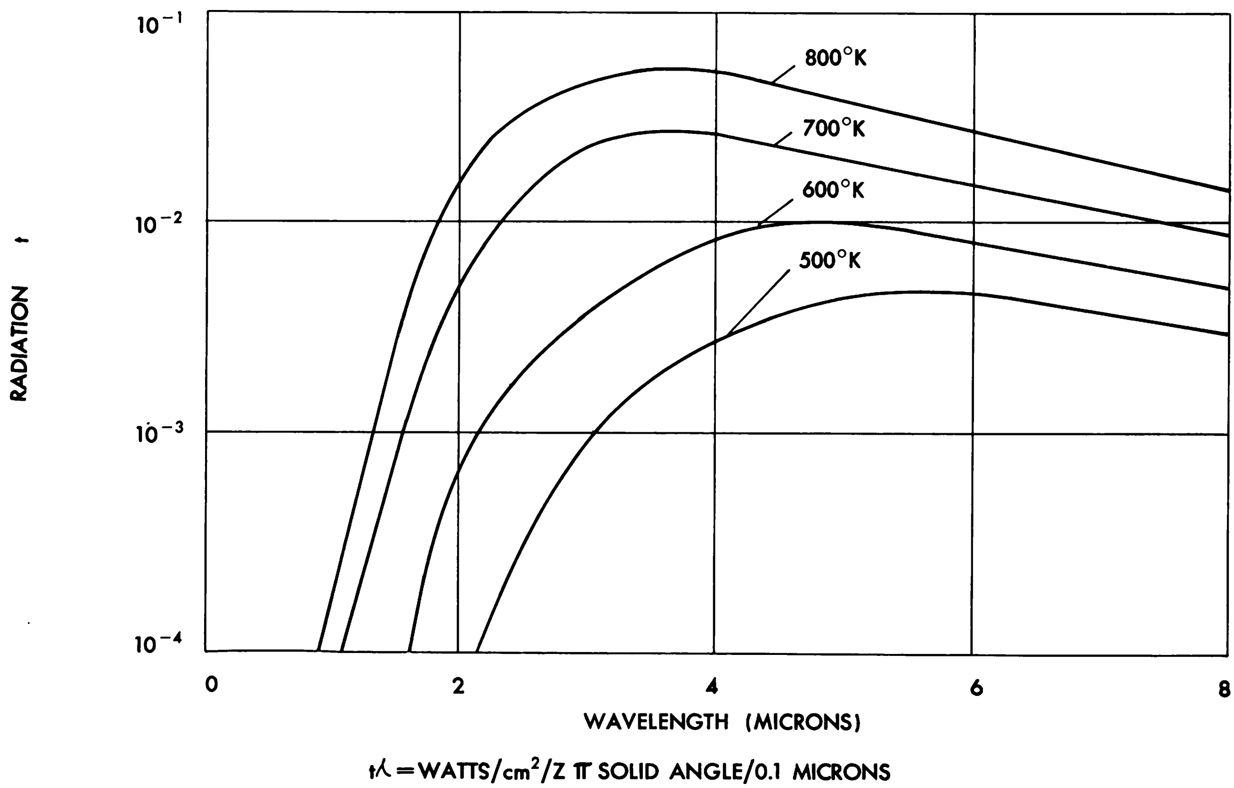
Characteristics of IR Radiators. An IR device senses *heat waves* or IR radiation emitted by the heated portions—main body and exhaust port—of target aircraft. These

so-called heat waves possess an amplitude and wavelength which are determined by the physical characteristics of the source such as the type of metal, the surface, the protective coatings, and the temperature. For example, if we assume two bodies, a carbon block and a polished aluminum block, are at an equal temperature, we would find that for any given wavelength the carbon body would emit infrared radiation of a greater amplitude than the aluminum block.

The illustration *Black Body Radiation Distribution as a Function of Temperature* shows the amplitude versus the wavelength or frequency of IR radiation from any body which we might assume to be an ideal or reference source. Two conclusions may be drawn from this illustration: (1) the peak relative amplitude occurs at increasing wavelengths as body or source temperature decreases; and (2) the relative amplitude of radiation increases as the source temperature increases. These two effects indicate an IR source is analogous to a radio signal containing a band of frequencies, with each possessing a certain amplitude.

In the case of an airplane with reciprocating engines, the radiation intensity of the exhaust gases decreases very rapidly as soon as the gases have left the exhaust manifold, due to the cooling effect of the atmosphere and the very fast intermixing of those gases with air. Hence, only the gases just behind the exhaust pipe exit need to be considered. This is not true for jet or rocket powered aircraft.

The illustration *Spectral Distribution of Radiation Emitted* represents, as an example, the spectral distribution of the radiation energy from the end of the exhaust pipe of a reciprocating airplane engine, including both the radiation of the end of the exhaust



Black Body Radiation Distribution as a Function of Temperature

pipe itself and the radiation of the exhaust gases immediately behind the pipe. The curve shows clearly that the spectral distribution is not equal to a black body radiation, but that some maxima and minima are present. Those maxima and minima are caused by emission and absorption of gases, especially water vapor and carbon dioxide, which are present in the exhaust gases. Both the spectral gas and metal radiation are overlapping each other.

In the case of the jet or rocket aircraft, the IR radiation from the exhaust closely approximates an ideal source radiator, as shown in the previously discussed illustration *Black Body Radiation Distribution as a Function of Temperature,* at a temperature of 700 degrees Kelvin. The main body or polished aluminum surface, though emitting IR radiation, actually resembles an ideal source radiator of a temperature less than the exhaust. It becomes evident that the aircraft simulates or is analogous to two transmitters: (1) polished aluminum, which possesses a low frequency and small amplitude; and (2) the jet exhaust, which possesses a relatively higher frequency and a greater amplitude.

The polished aluminum surface of the target is an isotropic (equal radiation in all directions) radiator of very low power as compared to radiation from the exhaust; present day detectors are not normally sensitive to this radiation because of the longer wavelength. Extending the analogy of the aircraft as an IR radiator to a radio transmitter, we may say the exhaust acts as a highly directional transmitter—the radiation pattern being proportional to β when β is defined as the angle off the source axis. Refer to diagram A of the illustration *Exhaust Radiation.*

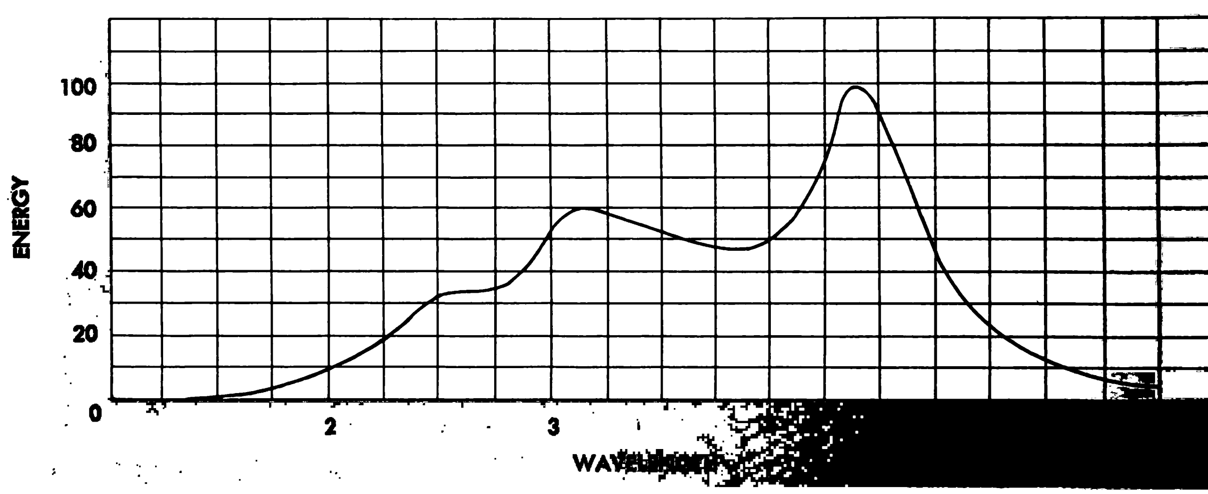
At distance R directly aft of the exhaust, the amount of radiation is J_1. It can be seen that the amount of radiation (J_2) at point X is

$$J_2 = J_1 \cos \beta$$

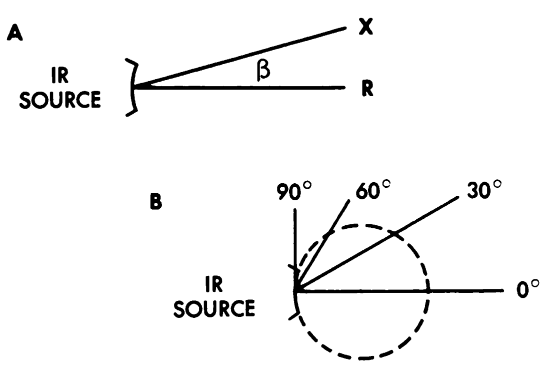
Varying angle β from zero to ninety degrees off the source axis (as shown in diagram B) will give a smaller radiation from the directional source.

Attenuation of Infrared Radiation

The IR radiation emitted by an airplane has to travel through the atmosphere before



Spectral Distribution of Radiation Emitted



Exhaust Radiation

it can be received by a device installed in another airplane. While traveling along its atmospheric path, the IR energy will be attenuated by absorption and scattering.

Absorption is mostly caused by water vapor and carbon dioxide in the atmosphere. Such an absorption takes place even in a clear atmosphere having an excellent visibility. The amount of absorption depends on the amount of water vapor and carbon dioxide within the air path through which the IR radiation has to travel. Since the water vapor and carbon dioxide contents decrease with an increase in altitude, the attenuation of IR radiation by absorption becomes smaller and smaller as the flight altitude increases.

Attenuation by scattering depends on the number and size of particles suspended in the atmosphere. These particles are mostly drops of condensed water vapor constituting haze, fog, or clouds. Such a cloudy atmosphere attenuates the IR radiation, and this is a great disadvantage of infrared methods. However, for many air-to-air applications the altitude of interest will be so great that neither absorption by water vapor or carbon dioxide or scattering by water particles will seriously decrease the detection ranges.

Homing Systems

Homing systems may be classified in three groups: passive, semiactive, and active. A passive homing system is one which is designed to select and identify the target by means of natural emanations or radiations from the target itself. Such radiations as heat waves, light waves, and sound waves have been used in passive homing systems.

A semiactive homing system is one which selects a target by means of energy from an external source, such as a tracking radar, reflecting from the target. This radar may be ground based or airborne. Equipment used in semiactive homing systems is more complex and bulky than that used in passive systems. It provides homing guidance over much greater ranges and with fewer external limitations in its application.

An active homing system is one in which the pilotless aircraft (P/A) carries its own source of energy with which to illuminate the target. The P/A homes on reflections or echoes of this energy from the target. This system is more complex than either the passive or semiactive types. However, it is well adapted for use in the terminal phase of long range P/A, as it is entirely independent of external sources of target illumination. Power supply requirements are the principal limiting factors in the use of active homing systems. At present this restricts their application to the larger, long-range P/A. We probably will find that as miniaturization of components is further developed, active homing systems will become more generally used.

INFRARED DETECTION

Infrared detectors may be likened to bandpass filters; that is, the sensing device is not only sensitive to IR amplitudes but also to the frequencies or the wavelengths of the radiated signals.

Components

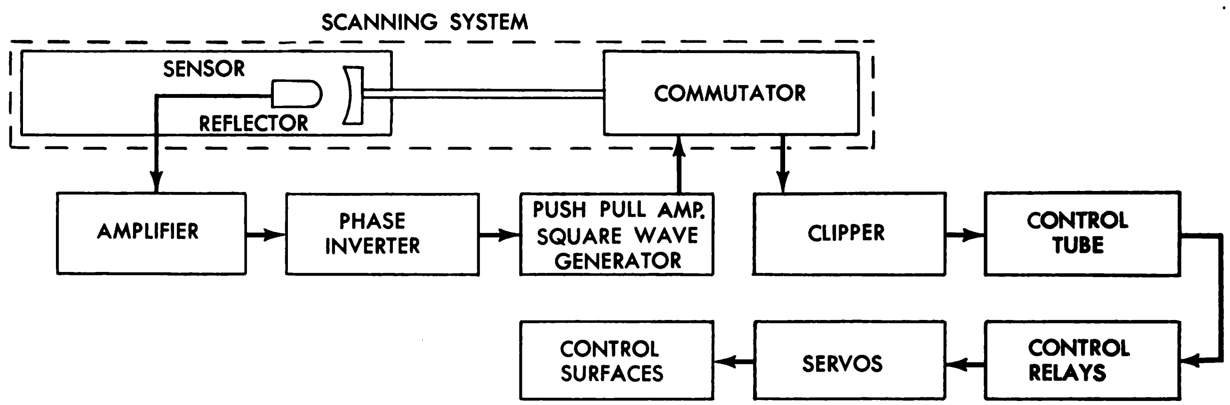
The basic components of an IR detection system include such items as a sensing element or optics, filters, a scanning system, reference generator, amplifiers, phase in-

verters, square wave generator, error signal computers, control systems, and relays. You are already familiar with most of these components.

The principal differences within passive homing systems lie in the type of sensing elements and scanning devices employed. The illustration *Passive Homing System* is representative of an IR detection system in general. The sensor which we use in such a system will be determined by the characteristic radiation from the target. These sensors are sometimes of the lead sulphide type or more often of the nickel-oxide type commonly known as bolometers.

pressure. The hydrogen gas tends to increase the heat absorbing property of the bolometer and causes it to respond more rapidly to slight variations in the amount of infrared radiation focused upon it. Bolometers possess the advantages of small size, rugged construction, and relatively large current-carrying capacity in addition to a high degree of sensitivity. For these reasons they have been used in heat seekers.

Heat radiations may cover a rather wide band of the light spectrum. So, it is essential that the sensing device be especially responsive to the radiations it is likely to receive throughout a wide range of atmospheric



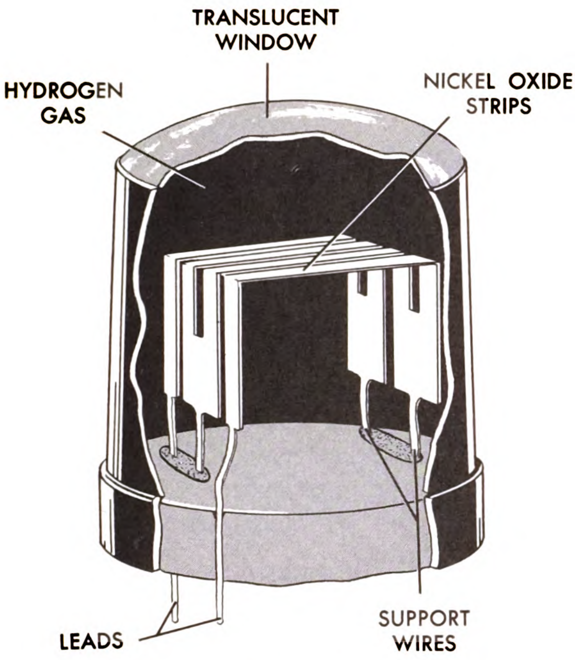
Passive Homing System

Within certain limits, lead sulphide possesses a high degree of *thermoresistivity* which is the property of a change in electrical resistance when subjected to changes in temperature. However, the lead sulphide cell is quite responsive to radiations in the light spectrum other than infrared; this fact reduces its usefulness as a light seeker.

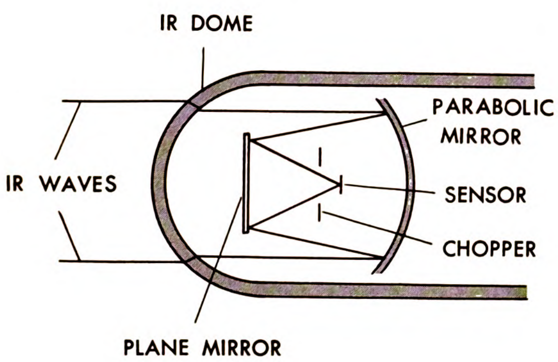
BOLOMETER. The *bolometer* (see illustration) has proved thus far to be the most efficient sensing unit for infrared homing systems. (This instrument has been used to measure the temperatures of stars.) It usually consists of very thin strips of black-coated (oxidized) nickel mounted in an evacuated cell or chamber into which has been injected a small amount of hydrogen gas under low

conditions, such as variations in humidity and dust concentration. Research has proved that infrared radiations of from eight to fourteen microns (eight to fourteen millionths of a meter) in wavelength comprise a large percentage of the total radiation of a body. This radiation is obstructed by moisture particles to a lesser degree than radiations of other wavelengths. Therefore, the bolometer is designed to be most sensitive to radiation within the eight to fourteen micron band.

OPTICAL SYSTEM. The illustration *Lens System* shows an example of a simple optical system that could be used in either an aircraft or a missile. IR energy radiating from a target or other source would pass through



Bolometer



Lens System

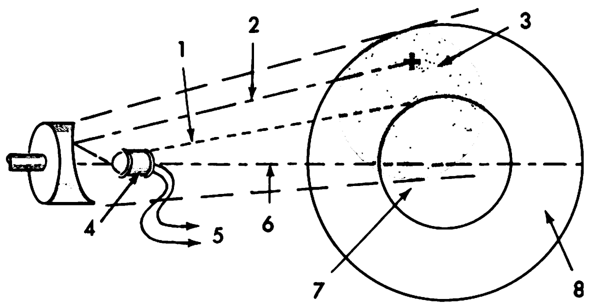
the IR dome and be collected on the parabolic mirror. Then it would be reflected back to the plane mirror and folded back to concentrate on the detector at the focal point. The motor-driven chopper would modulate the incoming energy, and the entire system would move about a vertical axis to follow the target.

LIGHT FILTERS. Very few materials absorb all wavelengths or colors of light uniformly. Therefore, when we have determined the wavelengths or colors of light which are desired, we can place a filter between the source of light and the sensor to absorb the undesirable wavelengths. Filters are generally made of colored glass, but they may be of other translucent materials possessing the desired qualities. For use with an IR sensor, we could select a filter which would absorb the violet, blue, and yellow light. This would permit the infrared light to pass through to the sensing element. We usually find that no single filter screen will produce the desired result; so two or more filters may be necessary to remove all or most of the undesired wavelengths. However, they must not reduce the intensity of the desired form of light below the sensitivity range of the sensing unit.

SCANNERS. By now you should be reasonably familiar with scanning systems as they operate on the same principle as a radar scanner. However, it is important for you to know how essential it is that the parabolic reflector surface or mirror used in scanners be of some material which will not absorb the radiations received from the target but will reflect them to the sensing element. Refer to the illustration *Conical Scanning System*. Notice how the sensor is located on the axis of rotation of the reflector. The entire scanning assembly should be well-balanced and free from mechanical and optical defects. Also, it should be mounted in a chamber to protect it from external pressure, moisture, and any other factors which might interfere with its operation. Other scanning methods applicable to homing systems include the fixed sector scanner, wherein several sensors are used, with each mounted or focused to view a specific portion of the total area scanned.

Basic Operation

A simple passive homing system could have an optical scanning system, an electronic amplifier, a demodulator, and a phase detector. The scanning system could be movable within a limited arc about its vertical



1. OPTICAL AXIS
2. RADIATION FROM TARGET
3. ROTATING FIELD OF VIEW
4. SENSOR
5. TO AMPLIFIER
6. AXIS OF ROTATION OF REFLECTOR
7. PATH OF ROTATING OPTICAL AXIS
8. TOTAL AREA SCANNED DURING ONE COMPLETE ROTATION

Conical Scanning System

and horizontal axes, and it could be gyro-stabilized.

Once locked on target, the missile is fired. The target deviations would then cause the optical head to turn and follow the target, generating an error voltage.

To acquire automatic angular tracking, conical scan could be used as it is used in radar. Thus, the viewing lobe as seen by the IR head would be made to describe a cone rotating at a fixed angular velocity. The images picked up by the parabolic disk would then appear to rotate in the focal plane of the optical system. The reference line for the angular tracking would be supplied by the central axis of the viewing lobe. Thus, any target on the axis would give a maximum amplitude signal. If the target moved off the axis, then the amplitude of the signal would decrease.

All the received IR signals would be optically modulated by a mechanical chopper just in front of the sensor unit. When focused on the sensor, these IR signals could then be transduced into electric signals varying at the chopping rate.

The chopped signal then would serve as a carrier that would be modulated in turn by the lobe signals (the envelope of the received signals that varies with the lobe rotation). After amplification, the lobe signal envelope would be demodulated from the composite signal.

The signal phase may then be compared with an internal phase reference sine wave (one that is synchronized with the lobe rotation). Any resultant phase difference, which is proportional to the angle the scanner is off target, would result in an error signal. This error signal is proportional in amplitude to the amount of displacement of the scanner from the central axis. The error signal would then be sent to a servo loop causing the scanner to turn and track the target.

When the scanner turns, it would cause the gyro to precess and generate a signal. This signal would then be sent to the control channel and steer the missile back on target and null the servo system.

To install search as well as track capabilities in the missile, a DC bias could be included in the system. Then when the input or received signal exceeded this bias (as when the receiver picks up radiation from a strong target), a relay could be activated changing the scanning from a search to a track scan.

To make the system effective, however, variables affecting the IR system, such as clouds, earth background, solar radiation, and abrupt transitions occurring at the horizon, would have to be overcome.

CHAPTER 9

HARMONIZATION

Harmonization is the procedure used to orient the guns, rockets, and missiles with the fire control system in an airplane. To achieve harmonization, the fire control system and armament components are correctly oriented by use of a harmonization target.

METHODS OF HARMONIZATION

There are two distinct methods of harmonization, the boresight and gunfire.

The Boresight Method

The boresight procedure is done on a 1000-inch range using a boresight target. The guns are aligned with the target by adjusting the rear gun mounts while sighting through the gun bore with a breech boresight tool. The sight and gun camera are then adjusted until aligned with the target index. The boresight method may be used after repair or replacement of guns, but only as a rough adjustment preceding the gunfire method. This procedure eliminates large corrections during the gunfire method.

On aircraft equipped with a nondirectional radar antenna, the antenna must be adjusted by this method.

The Gunfire Method

The gunfire method may take place on firing ranges that vary from 1000 to 2250 feet in length. This range must have a tie-down slab placed at the selected distance from the firing abutment. The target which is placed in the firing abutment is positioned in correct relationship to the secured airplane. The sight and gun camera are adjusted to align the sight line and camera view to the target index. The guns are then fired and adjusted until correctly aligned to get the maximum number of hits in the desired dispersion area as shown on the target.

Accurate harmonization requires the same combination of gunfire that is used during aircraft flight. This is necessary because the dispersion relative to a gun, when fired individually, may be changed when all the guns are fired. This action is referred to as rhythm shift.

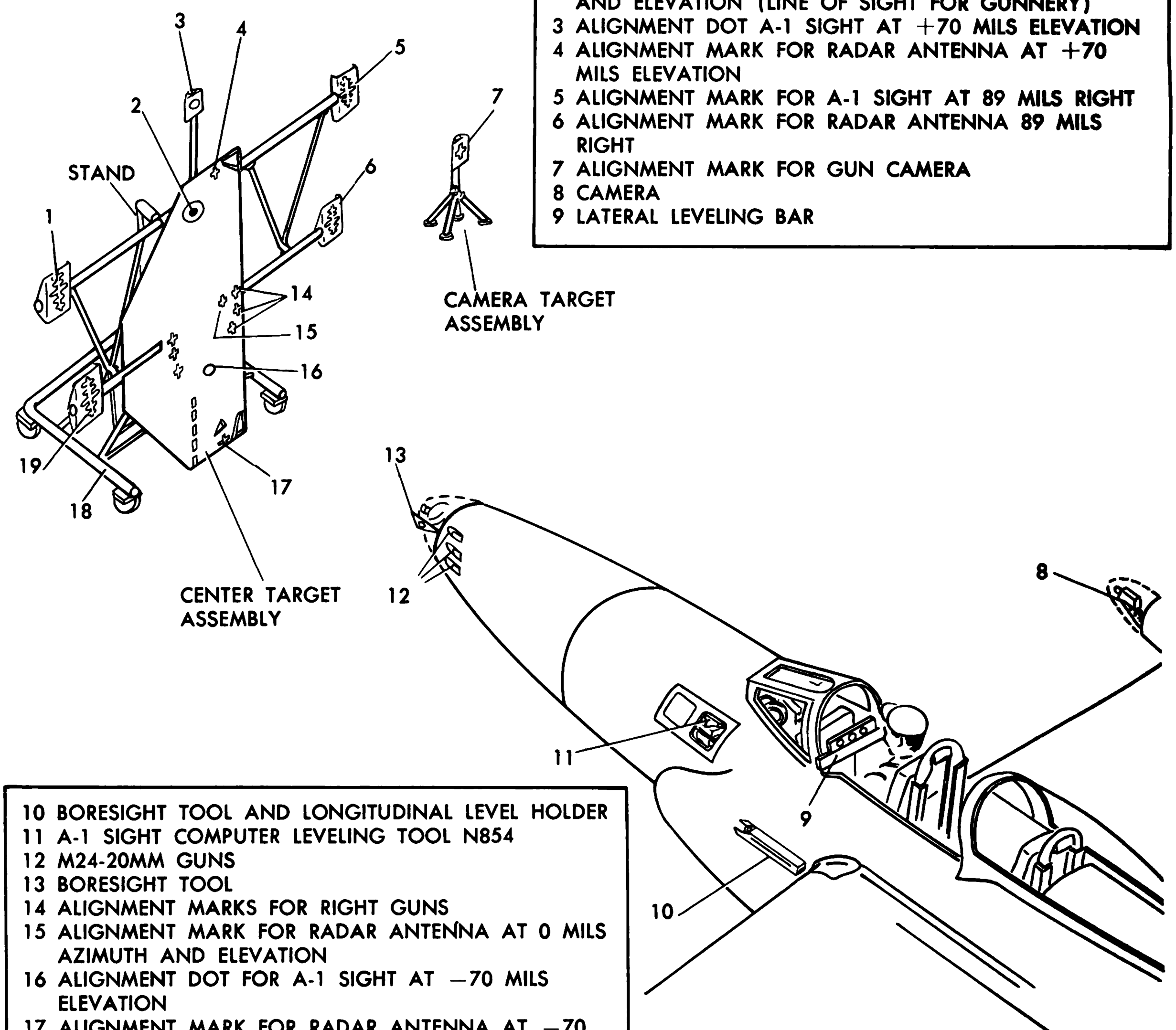
Gun Configuration

The guns may be adjusted to produce a guns-parallel type of gunfire (projectile paths do not cross) or a guns-converged type of gunfire (projectile paths intersect at a point). The parallel guns configuration allows a lethal density of projectiles for all ranges. Converged guns allow a maximum density for one range only.

Parallax Correction Range

The sight in the F-100D aircraft is about 58 inches above the mean gun line, thus introducing parallax. This is compensated for

1 ALIGNMENT MARK FOR A-1 SIGHT FOR 89 MILS LEFT
2 ALIGNMENT MARK FOR A-1 SIGHT AT 0 MILS AZIMUTH AND ELEVATION (LINE OF SIGHT FOR GUNNERY)
3 ALIGNMENT DOT A-1 SIGHT AT +70 MILS ELEVATION
4 ALIGNMENT MARK FOR RADAR ANTENNA AT +70 MILS ELEVATION
5 ALIGNMENT MARK FOR A-1 SIGHT AT 89 MILS RIGHT
6 ALIGNMENT MARK FOR RADAR ANTENNA 89 MILS RIGHT
7 ALIGNMENT MARK FOR GUN CAMERA
8 CAMERA
9 LATERAL LEVELING BAR



10 BORESIGHT TOOL AND LONGITUDINAL LEVEL HOLDER
11 A-1 SIGHT COMPUTER LEVELING TOOL N854
12 M24-20MM GUNS
13 BORESIGHT TOOL
14 ALIGNMENT MARKS FOR RIGHT GUNS
15 ALIGNMENT MARK FOR RADAR ANTENNA AT 0 MILS AZIMUTH AND ELEVATION
16 ALIGNMENT DOT FOR A-1 SIGHT AT −70 MILS ELEVATION
17 ALIGNMENT MARK FOR RADAR ANTENNA AT −70 MILS ELEVATION
18 ALIGNMENT MARKS FOR LEFT GUNS
19 ALIGNMENT MARKS FOR RADAR ANTENNA 89 MILS LEFT

Boresighting

during harmonization by adjusting the caged sight line to intersect the mean gun line at a given distance from the aircraft. This distance, called parallax correction range, is 2250 feet for parallax gun configuration (Air Force specifications). If gunfire is to be converged, the sight may be adjusted for parallax correction at the convergence range instead of 2250 feet. Parallax ranges below 1000 feet should not be used. The maximum practical airplane-to-target distance should be used for harmonization to gain the maximum accuracy in aligning the armament components.

Range for the Gunfire Method

Range selection is usually limited in both location and length by conditions such as airfield boundaries, runways, and inhabited areas. The most accurate results are obtained using the expected combat range. Ranges of 1000 feet and 1800 feet are the most common. Ranges shorter than 1000 feet should not be used.

An aircraft tie-down slab must be constructed at the desired position on the range. This slab must have tie-down rings and markers so that the aircraft may be properly positioned and tied down in relation to the target. Tie-down of the aircraft serves two purposes during gunfire: It prevents aircraft movement with reference to the target, and keeps the aircraft from being displaced from the jacks.

Safety precautions must be observed at all times when firing-in an aircraft. All guns must be cleared before the pit crew checks the target for hits. A field telephone or some other means of communications must be used to keep the firing-in crew in touch with the pit crew. Gunfire range warning devices such as red flags and caution signs will be used.

Target Selection

The proper selection of the target is a major factor for correct harmonization. The target provides a visible record of the size and placement of the actual pattern of each gun after firing. The guns are fired and adjusted until the actual dispersion falls within the desired area on the target. The sight is adjusted until the tracking index (pipper) of the reticle image coincides with the respective target cross. The gun camera is then adjusted until its view is properly oriented. The guns, sight, and gun camera all depend on the target for correct adjustment. Care should be taken during target construction. Failure to do so will result in inaccurate harmonization.

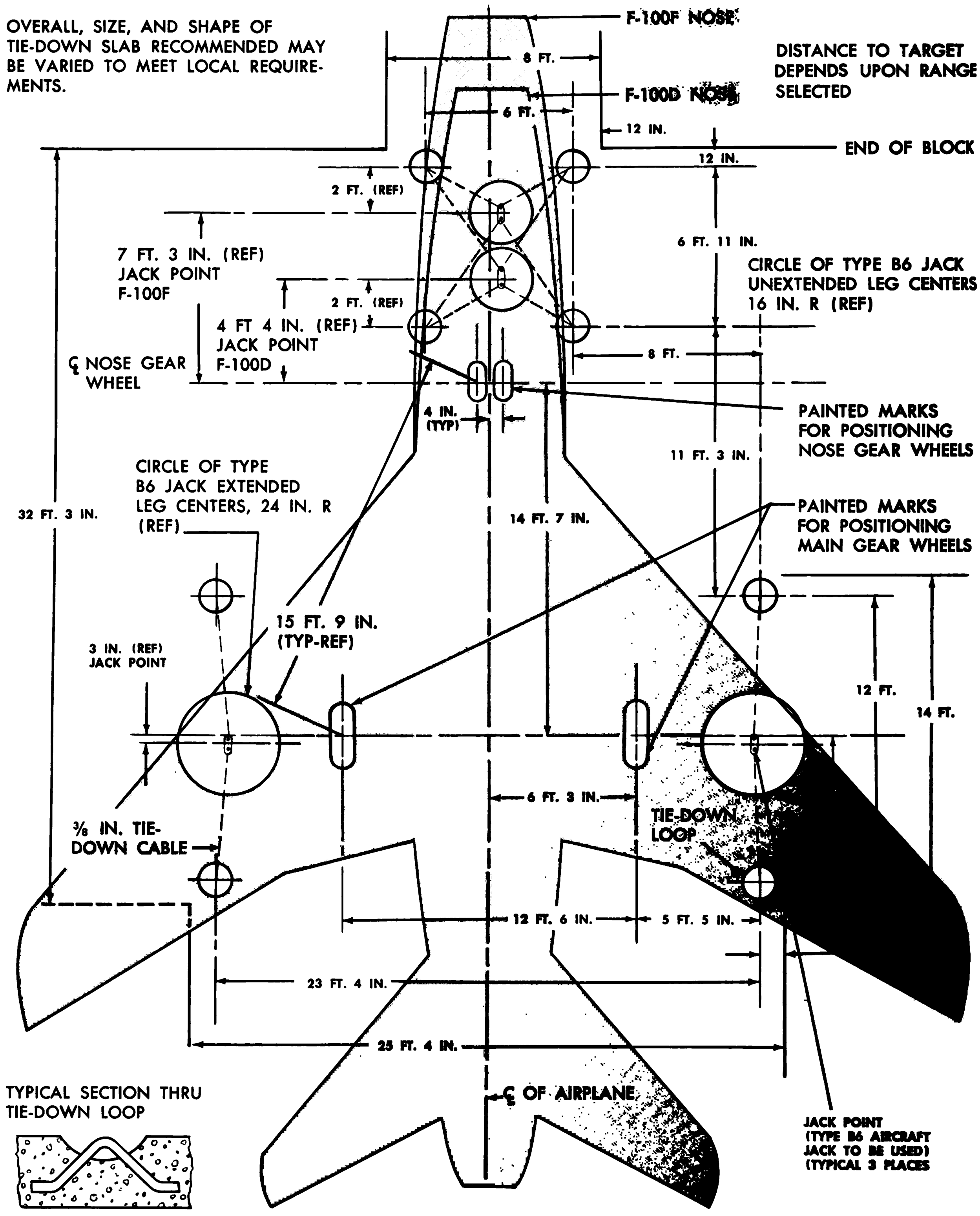
Two factors determine the target selected for gunfire methods of armament harmonization. The first factor is the distance between aircraft and target. This is normally predetermined by the length of the available range. The other factor is the gun configuration to be used (guns parallel or guns converged).

Once the proper target has been selected, it can be constructed more easily by using a precut stencil. If the guns are very far off, it may take more than one target to correctly harmonize the aircraft.

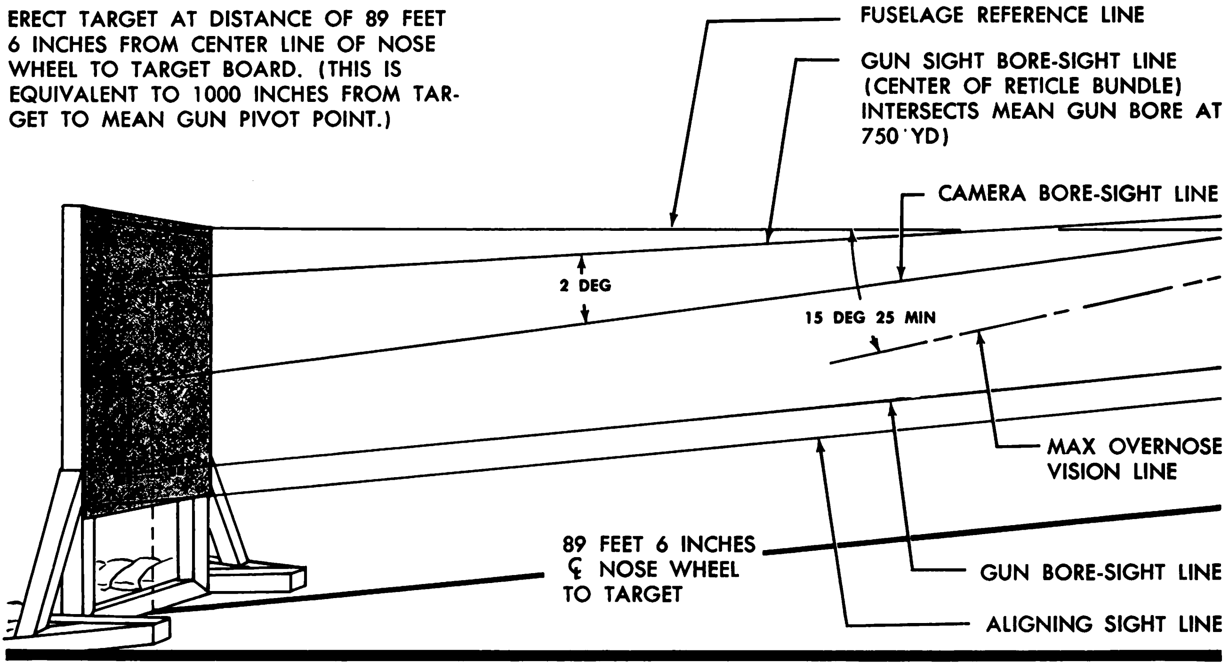
RADIO-FREQUENCY RADIATION HAZARDS

When firing-in the aircraft, be sure the radar system is turned off. All Air Force radio-frequency (RF) equipment is designed and installed so that the danger to operating personnel from exposure to RF energy, including microwave radiation during normal operation, is minimized. There are certain actions that should be avoided by maintenance personnel and others during operating and maintenance periods. Radio-frequency radiation may cause damage to human tissues if it is encountered in sufficient quantity. Exposure of four to five minutes to a microwave radiation intensity of 0.2 watts/cm^2 has produced cataracts of the eyes in experimental animals. Research is in progress to determine the biological effects of radiation on personnel. Until these studies are complete, all RF radiation must be considered as dangerous. Precautions should be taken to avoid unnecessary exposure of personnel to radiation energies. To avoid unnecessary hazards to personnel, the following precautionary measures should be taken.

All areas in which an individual may be exposed to radio-frequency energy having a power density of 0.01 watts/cm^2 should be considered danger areas. All hazardous areas should have appropriately posted warning signs, and personnel should not be permitted in such areas while the radar is operating. The practice of discharging under test the RF output of high-power generators that produce power levels of 0.01 watts/cm^2 or

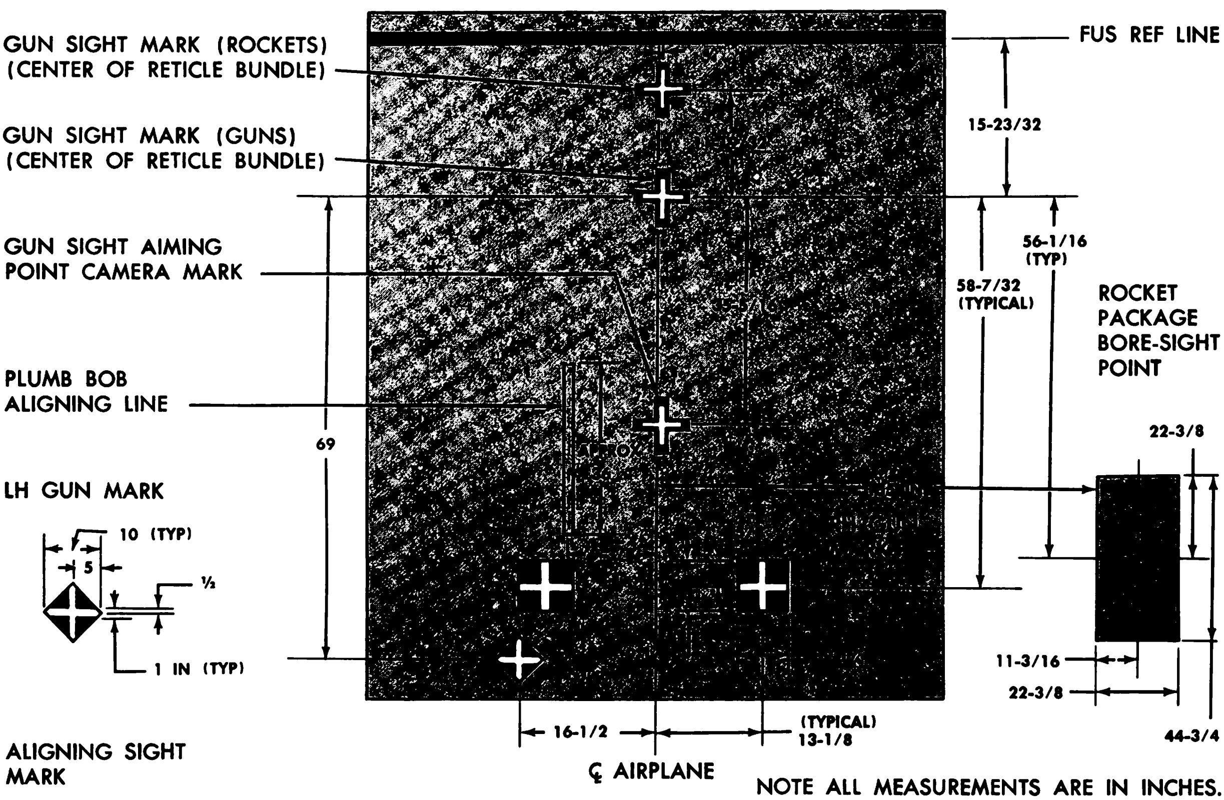


Tie-Down Slab for Harmonization of F-100D Aircraft



ALL DIMENSIONS ARE TRUE

TARGET CONSTRUCTION



Typical Target Construction for F-100D Aircraft

more into the surrounding areas must be discouraged.

Dummy loads, water loads, or other absorptive materials should be used to absorb the energy output of such equipment being operated or tested.

Where test procedures require free-space radiation, the radiating device should be located to avoid directing the energy beam toward inhabited structures or other personnel groups. In the positioning of such radiating devices, care must be taken to avoid reflecting either the primary beam or accessory lobes in such a manner as to expose personnel in adjacent areas.

Personnel should be prohibited from performing any work on antennas or waveguides while the equipment is in operation.

HARMONIZATION (BORESIGHTING) OF F-101A AIRCRAFT

Harmonization Equipment

The equipment for aircraft harmonization will consist of the following items: nose and wing jacks to level the aircraft, plumb bombs which are used to correctly align the target with the aircraft, a spirit level and leveling bars to level the aircraft. You must also have a ground power unit to supply the aircraft with necessary power.

Boresighting the Guns

With the target in place, the aircraft nose should be raised 11 inches. The main landing gear wheels must be chocked. Check alignment of boresight target index by sighting through the boresight fixtures of the aircraft.

1. Disconnect feed chutes from all four guns.
2. Install special boresight tool adapter (T496) into gun bore.
3. Install A.F. caliber .50 boresight tool from the J-1 Kit into the adapter.
4. Sight through the boresight tool and align the gun with the boresight target by alternately turning the azimuth and elevation adjusting screws on rear gun mount. Repeat the procedure for the other three guns.

Boresighting the Gunsight

The MA-7 sight is boresighted internally at the time of manufacture; therefore, no further adjustments are necessary except adjusting of the boresight mirrors by means of two external adjustments. The adjustments have screwdriver slots; they are located at the right and above the collimating lens.

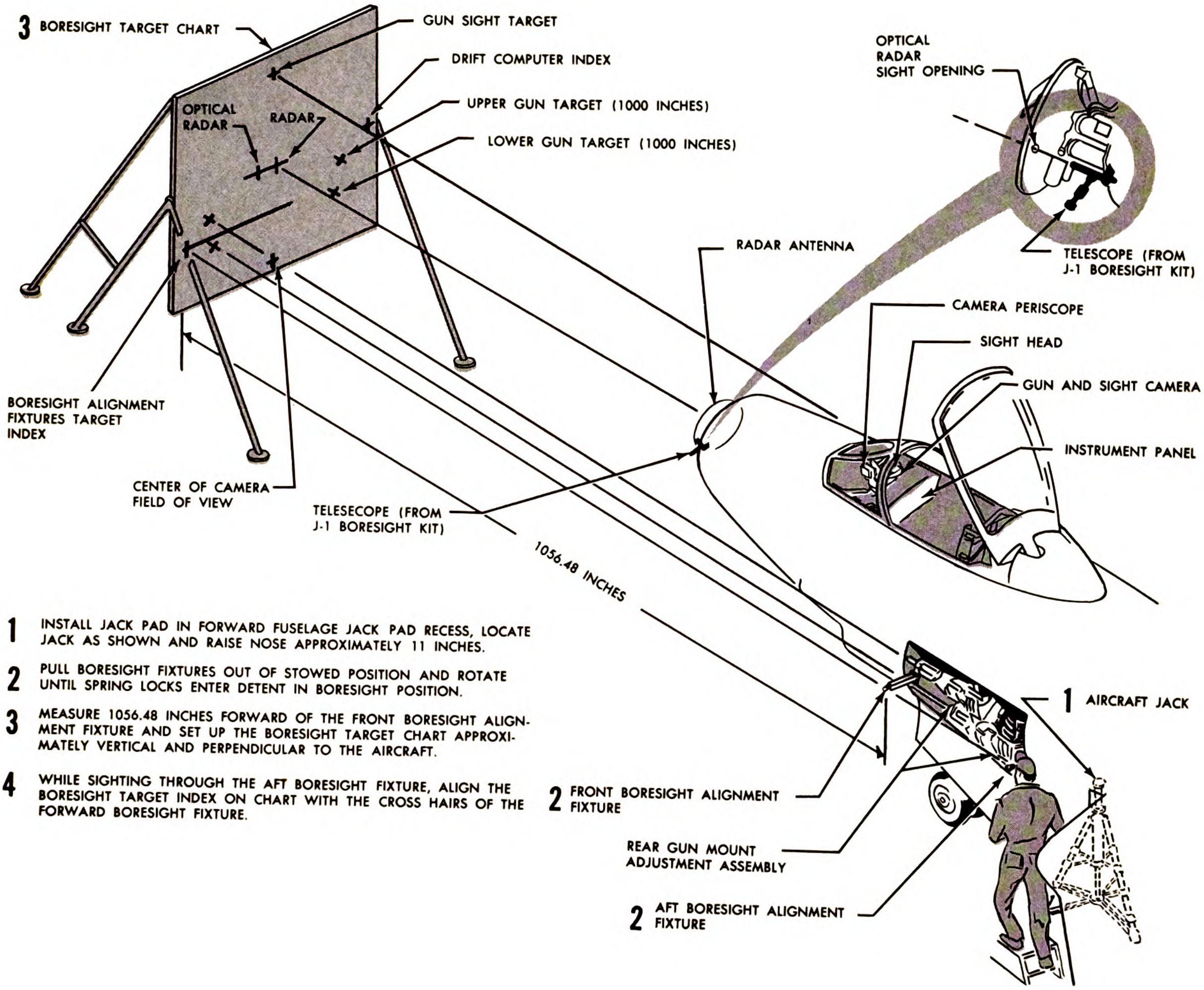
With the external power connected to the aircraft and the sight system turned on, the gyroscope must be mechanically caged. This is accomplished by rotating the cage knob to the left until stopped by the mechanical stop in the sight head.

Center the reticle as closely as possible on the target by rotating the sight head and yoke assembly on the serrated base for azimuth alignment. Make the elevation adjustment by use of the threaded post between the yoke assembly and sight base. Torque the AN6-21A clamp bolt, that locks the halves of the serrated base together, to 160-180 inch-pounds.

The fine adjustments are made by rotating the boresight adjustments. Total fine adjustment of the boresight mirrors is 10 mils in either azimuth or elevation. If greater adjustment is necessary, readjust the sight head on its mounts to bring it within range of the limits of the fine adjustments.

Camera Periscope Boresighting

The camera must be removed from the sight head. With the gunsight mechanically caged and external electrical power ON, set the gun master switch to SIGHT-CAMERA. Rotate the manual range grip to adjust the reticle image to its maximum diameter.

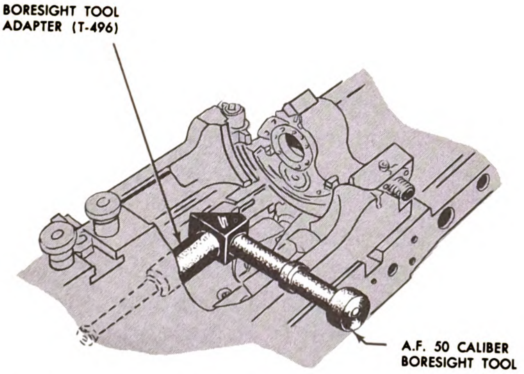


1 INSTALL JACK PAD IN FORWARD FUSELAGE JACK PAD RECESS, LOCATE JACK AS SHOWN AND RAISE NOSE APPROXIMATELY 11 INCHES.

2 PULL BORESIGHT FIXTURES OUT OF STOWED POSITION AND ROTATE UNTIL SPRING LOCKS ENTER DETENT IN BORESIGHT POSITION.

3 MEASURE 1056.48 INCHES FORWARD OF THE FRONT BORESIGHT ALIGNMENT FIXTURE AND SET UP THE BORESIGHT TARGET CHART APPROXIMATELY VERTICAL AND PERPENDICULAR TO THE AIRCRAFT.

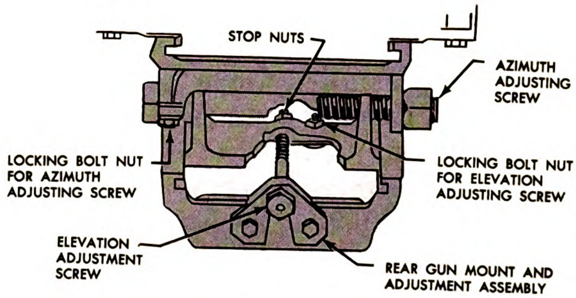
4 WHILE SIGHTING THROUGH THE AFT BORESIGHT FIXTURE, ALIGN THE BORESIGHT TARGET INDEX ON CHART WITH THE CROSS HAIRS OF THE FORWARD BORESIGHT FIXTURE.



5 REMOVE THE GUN FEEDER AND THE DRUM ASSEMBLY FROM EACH GUN (REFER TO "M-39 GUNS REMOVAL".) INSTALL BORESIGHT TOOL ADAPTED (T-496) IN THE BREECH. SLIDE A.F. 50 CALIBER BORESIGHT TOOL (FROM J-1 KIT) ONTO ADAPTER (T-496).

NOTE

TORQUE REQUIRED TO TURN AZIMUTH AND ELEVATION SCREWS MUST BE 75 ± 15 INCH-LBS. TO OBTAIN SPECIFIC TORQUE, TIGHTEN LOCKING BOLTS ON RESPECTIVE SCREWS.



6 WHILE SIGHTING THROUGH THE BORESIGHT TOOL, ALTERNATELY ADJUST THE AZIMUTH AND ELEVATION SCREWS UNTIL THE GUN IS CENTERED ON THE TARGET.

Boresight Method F-101A Aircraft

With the aiming dot of the gunsight reticle aligned with the sight target index, align the camera periscope so that the reticle is centered on the same boresight target index.

Camera Boresighting

The camera must be mounted on the sight head. Remove the magazine from the camera.

1. Mount the A-2 camera boresighting tool on the camera.
2. Loosen the screws on the camera mounting bracket. The screw holes are elongated to permit movement for alignment.
3. Set the camera diaphragm opening to f/2.8. While viewing through the boresight tool, simultaneously depress and turn the manual drive knob on the aft end of the camera until the shutter blades in the camera assembly are moved out of the line of sight.
4. Focus the boresight tool by adjusting the eyepiece in or out until the crosshairs present a sharply defined image.
5. While viewing the reticle through the boresight tool, move the camera to center the aiming dot transversely and near the top of the camera field. Tighten the mounting bracket screws.

HARMONIZATION (GUNFIRE METHOD) OF F-100 AIRCRAFT

After the aircraft is towed to the tie-down slab for harmonization, it must be properly jacked, leveled laterally, and tied down to the slab. The aircraft must be leveled laterally to establish a desired relationship between the guns in the aircraft and the target. The nose of the aircraft can be raised or lowered and the target moved left or right so that the cross on the ring sight, seen through the peep sight, is superimposed on the aligning sight mark on the target.

WARNING

Make sure that all guns are clear of ammunition.

1. Position aircraft on tie-down slab.
2. Using E4214 gun harmonization and tie-down kit, install E3136 jack adapter on fuselage forward section, and install 3137-1 and -2 jack adaptors on left and right wings. Torque the mounting studs to 40-50 inch-pounds with a torque wrench. Tighten down locknuts to 30-40 inch-pounds.

NOTE

The locknuts must not be tightened down and torqued individually. They should be tightened gradually to have even pressure on the jack pad at each mounting stud.

3. Place a 10-ton or greater capacity jack under each wing.
4. Raise the jacks at the same time until the landing gear struts begin to extend and the greater part of the aircraft weight is on the jacks.
5. Level the aircraft laterally. Place a long level or a level protractor and level bar on the leveling lugs on the aft bulkhead in the right main landing gear wheel well. Raise or lower main landing gear jacks until the aircraft is laterally level.
6. Position selected target in firing abutment.
7. Install ring and peep sight fixtures (T448-1 kit). The peep sight is positioned on the forward face of the

left main wheel well and the ring sight is positioned on the forward face of the left gun bay.

8. While sighting through ring and peep sight, raise or lower nose wheel jack to align aircraft with target alignment sight.

NOTE

Do not disturb the lateral level of the aircraft. It may be necessary to move the target to the left or right until crosshairs of the ring sight fixture are superimposed on the aligning sight mark on target. A white border of uniform width will appear around crosshairs of the ring sight fixture when sighting through the ring sight.

9. Install tie-down cables through the jack pads to the slab tie-down rings, and adjust tension cables to prevent airplane displacement on the jacks during gunfire.

NOTE

For accurate harmonization results, the aircraft and target must be properly secured to eliminate all aircraft and target motion. This condition must be checked at short intervals throughout the procedure by sighting through the ring and peep sight fixtures to the target aligning sight mark.

Range Safety Precautions

Extreme caution must be exercised by authorized personnel to make sure that personnel, ground equipment, and installations are clear of gunfire range.

Direct communication should be kept between target area and aircraft. Firing will begin only after the target crew has cleared the firing abutment area.

The flight control tower should be notified that gunfire harmonization is in progress. A visual warning sign (flag or sign) should be posted.

After each firing mission or after a full complement has been fired, drums of both guns must be removed for visual inspection of all drum seals and breech face of the barrels. If the seals are found cracked, they must be replaced. Particular attention must be given to the breech face of the barrel. Replace the barrel if there is any evidence of erosion or breech face washout.

To eliminate firing out of battery position, remove gun-firing harness assembly from the gun after 1000 rounds have been fired. Wash the electrical knife-blade contacts and knife blade with plain water and dry all items thoroughly before installing.

NOTE

Each gun should be fired and adjusted for the correct dispersion pattern. After each gun has been correctly adjusted, then fire both guns in a salvo to check firing as in actual flight conditions.

INDEX

A

B

C

D

E

F

G

H

M

N

O

S

T

V

W

Y

Z

☆ U.S. GOVERNMENT PRINTING OFFICE: 1959—520985/20